

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 2, 2001, 00:02:02 ; Search time 66.17 Seconds  
(without alignments)  
372.057 Million cell updates/sec

Title: US-09-331-723A-4  
 Perfect score: 141  
 Sequence: 1 qcgcgcgagccctggctc.....agctgcaccgcgcacgcag 141

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 302621 seqs, 87307344 residues

Total number of hits satisfying chosen parameters: 605242

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

```
Post-processing: Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 s
```

```
Database : Issued_Patents_NA:*
1: /cgn2_6/pdata1/1/na/5A_COMB.seq:*
2: /cgn2_6/pdata1/1/na/5B_COMB.seq:*
3: /cgn2_6/pdata1/1/na/6A_COMB.seq:*
4: /cgn2_6/pdata1/1/na/6B_COMB.seq:*
5: /cgn2_6/pdata1/1/na/PCRTUS_COMB.seq:*
6: /cgn2_6/pdata1/1/na/backfiles1.seq:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Query			Length	DB	ID	Description
	Score	Match					
C 1	139.4	98.9	3381	5	US-09-009-119-1		Sequence 1, Appli
C 2	136.2	96.6	3383	4	PCR-US95-09098-1		Sequence 1, Appli
C 3	38.8	27.5	2900	1	US-08-034-650-9		Sequence 9, Appli
C 4	38.8	27.5	2900	1	US-08-449-015-9		Sequence 9, Appli
C 5	38.2	27.1	30001	1	US-08-135-168-1		Sequence 1, Appli
C 6	38.2	27.1	30001	2	US-08-474-933-1		Sequence 1, Appli
C 7	38	27.0	2310	4	US-09-036-987A-25		Sequence 25, Appli
C 8	34.2	24.3	43280	2	US-08-804-227C-1		Sequence 1, Appli
C 9	34	24.1	1485	1	US-07-702-771-2		Sequence 1, Appli
C 10	34	24.1	1485	1	US-08-438-883-2		Sequence 2, Appli
C 11	34	24.1	1485	1	US-08-741-599-2		Sequence 2, Appli
C 12	34	24.1	2685	2	US-08-935-450-5		Sequence 5, Appli
C 13	33.8	24.0	1212	2	US-08-449-933-5		Sequence 5, Appli
C 14	33.8	24.0	8937	2	US-08-449-933-1		Sequence 1, Appli
C 15	33.8	24.0	10706	1	US-08-411-389-1		Sequence 1, Appli
C 16	33.4	23.7	2634	3	US-08-911-853-30		Sequence 30, Appli
C 17	33.4	23.7	17612	3	US-08-911-853-29		Sequence 29, Appli
C 18	33.2	23.5	1215	3	US-08-911-853-8		Sequence 8, Appli
C 19	33.2	23.5	17612	3	US-08-911-853-29		Sequence 29, Appli
C 20	32.8	23.3	1032	1	US-08-400-422-2		Sequence 2, Appli
C 21	32.8	23.3	1951	3	US-08-922-865-1		Sequence 1, Appli
C 22	32.6	23.1	862	1	US-08-379-280-4		Sequence 4, Appli
C 23	31.8	22.6	4649	6	5183745-1		Patent No. 5183745
C 24	31.8	22.6	6443	6	5193745-5		Patent No. 5193745
C 25	31.8	22.6	20235	1	US-07-642-734C-3		Sequence 3, Appli
C 26	31.8	22.6	20235	3	US-08-439-009A-3		Sequence 3, Appli
C 27	31.6	22.4	900	5	PCR-US95-04801-3		Sequence 3, Appli

## ALIGNMENTS

```

RESULT 1
US-09-009-119-1/c
; Sequence 1, Application US/090091119
; Patent No. 6160206
; GENERAL INFORMATION:
; APPLICANT: SATO, RYO
; APPLICANT: BOYNTON, John E.
; APPLICANT: BOYNTON, John E.
; APPLICANT: GILLHAM, Nicholas W.
; APPLICANT: HARRIS, Elizabeth H.
; TITLE OF INVENTION: Porphyrin Accumulating-Type Herbicide Resistance Gene
; FILE REFERENCE: substitute sequence listing
; CURRENT APPLICATION NUMBER: US/09/009,119
; CURRENT FILING DATE: 1998-01-20
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 3381
; TYPE: DNA
; ORGANISM: Chlamydomonas reinhardtii
US-09-009-119-1

Query Match 98.9%; Score 139.4; DB 4; Length 3381;
Best Local Similarity 99.3%; Pred. No. 7.1e-25;
Matches 140; Conservative 0; Mismatches 1; Indels 0; Gaps

QY 1 gccgcgagccctggactcttcgactaccgcgcggtgggcgcctgacactatcgatc 60
Ddb 3127 gccgcgcgagccctgggctccttcgactaccgcgcgcggtgggcgcctgacactatcgatc 60
QY 61 ccgctgagccgctgcgggaggagcgcaaggcctcgacgggtccgtgcggggttcgg 120
Ddb 3067 ccgctgagccgctgcgggaggagcgcaaggcctcgacgggtccgtgcggggttcgg 120
QY 121 cagctgcaccgcgcgcagcag 141
Ddb 3007 cagctgcaccgcgcgcagcag 141

RESULT 2
PCT-US95-09098-1/c
; Sequence 1, Application PC/TUS9509098
; GENERAL INFORMATION:
; APPLICANT: SATO, RYO
; APPLICANT: Boynton, John E.
; APPLICANT: Gillham, Nicholas W.
; APPLICANT: Harris, Elizabeth H.
; TITLE OF INVENTION: Porphyrin-Accumulating Type Herbicide
; TITLE OF INVENTION: Resistance Gene
; NUMBER OF SEQUENCES: 1

```

CORRESPONDENCE ADDRESS:  
ADDRESSEE: Birch, Stewart, Kolasch & Birch, LLP  
STREET: 8110 Gatehouse Road, Suite 500 East  
CITY: Falls Church  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22042  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/09098  
FILING DATE: 20-JUL-1995  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Murphy Jr., Gerald M.  
REGISTRATION NUMBER: 28,977  
REFERENCE/DOCKET NUMBER: 2185-110P  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 205-8000  
TELEFAX: (703) 205-8050  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3383 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Chlamydomonas reinhardtii  
STRAIN: RS-3  
PCT-US95-09098-1

Query Match 96.68; Score 136.2; DB 5; Length 3383;  
Best Local Similarity 97.98; Pred. No. 3.9e-24;  
Matches 138; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1 gccgcgagccctggctcttcgactaccgcgcggtggcgccgtgaagctgtgtac 60  
|||||  
DB 3128 gccgcgagccctggctcttcgactaccgcgcggtggcgccgtgaagctgtgtac 3069  
QY 61 ccgctagccgctgagagagcgcgaagcctcgagcgggtcgtgcccggcttcggt 120  
|||||  
DB 3068 ccgctagccgctgagagagcgcgaagcctcgagcgggtcgtgcccggcttcggt 3009  
QY 121 cagctgacccgcgcacgcag 141  
|||||  
DB 3008 catctgacccgcgcacgcag 2988

RESULT 3  
US-08-034-650-9  
; Sequence 9, Application US/08034650  
; Patent No. 5641671  
; GENERAL INFORMATION:  
; APPLICANT: BOS, Jannetje W.  
; APPLICANT: FRENKEN, Leon G.  
; APPLICANT: VERRIPS, Cornelis T.  
; APPLICANT: VISSER, Christiaan  
; TITLE OF INVENTION: PRODUCTION OF ACTIVE PSEUDOMONAS GLUMAE  
; TITLE OF INVENTION: LIPASE IN HOMOLOGOUS OR HETEROLOGOUS HOSTS  
; NUMBER OF SEQUENCES: 13  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CUSHMAN, DARBY & CUSHMAN  
; STREET: 1615 L. Street, N.W.  
; CITY: Washington  
; STATE: D.C.

COUNTRY: USA  
ZIP: 20036-5601  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/034,650  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/727,235  
FILING DATE: 03-JUL-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Kokulis, Paul N.  
REGISTRATION NUMBER: 16,773  
REFERENCE/DOCKET NUMBER: PNK/5970/91731  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 861-3000  
TELEFAX: (202) 822-0944  
TELEX: 6714627 CUSH  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2900 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 483..1556  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1559..2617  
US-08-034-650-9

Query Match 27.58; Score 38.8; DB 1; Length 2900;  
Best Local Similarity 55.18; Pred. No. 0.15;  
Matches 76; Conservative 0; Mismatches 62; Indels 0; Gaps 0;  
QY 2 ccgcgcgagccctggctcttcgactaccgcgcggtggcgccgtgacgtgtcgctacc 61  
|||||  
DB 2046 ccgcgcgagccctggctcttcgactaccgcgcggtggcgccgtgacgtgtcgctacc 2105  
QY 62 cgcgtgagcgcgtgagagagcgcgaagcctcgagcgggttcgtgcccgggttcggtc 121  
|||||  
DB 2106 atcagccgcgcgcgtgagagcgcgaagcctcgagcgggttcgtgcccgggttcggtc 2165  
QY 122 agctgacccgcgcacgc 139  
|||||  
DB 2166 acgagcagccgcgcacgc 2183

RESULT 4  
US-08-449-015-9  
; Sequence 9, Application US/08449015  
; Patent No. 5804409  
; GENERAL INFORMATION:  
; APPLICANT: BOS, Jannetje W.  
; APPLICANT: FRENKEN, Leon G.  
; APPLICANT: VERRIPS, Cornelis T.  
; APPLICANT: VISSER, Christiaan  
; TITLE OF INVENTION: PRODUCTION OF ACTIVE PSEUDOMONAS GLUMAE  
; TITLE OF INVENTION: LIPASE IN HOMOLOGOUS OR HETEROLOGOUS HOSTS  
; NUMBER OF SEQUENCES: 13  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CUSHMAN, DARBY & CUSHMAN  
; STREET: 1615 L. Street, N.W.  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20036-5601

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/449,015  
FILING DATE: 24-MAY-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/727,235  
FILING DATE: 03-JUL-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Kokulis, Paul N.  
REGISTRATION NUMBER: 16,773  
REFERENCE/DOCKET NUMBER: PNK/5970/91731  
TELEPHONE: (202) 861-3000  
TELEFAX: (202) 822-0944  
TELEX: 6714627 CUSH  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2900 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 483..1556  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1559..2617  
US-08-449-015-9

Query Match 27.5%; Score 38.8; DB 1; Length 2900;  
Best Local Similarity 55.1%; Pred. No. 0.15;  
Matches 76; Conservative 0; Mismatches 62; Indels 0; Gaps 0;  
QY 2 ccgcgagcccttgctcttcactaccgcgcggtgagcgccgtgacgtgtgtacc 61  
Db 2046 CGCGGACGGCGGTGCTCGCGGACAGCTGATCGCGCGCGCATGACGCTCGCGCTCG 2105  
QY 62 cgtgagccgctgctgagcgagcgcaaggcctcgacggttcgtgcgggcttcggtc 121  
Db 2106 ATACGCGCGCGCGCTGCGCGACCGACGCTCGCGGAGTGGCGGCGGTCTTTCGCG 2165  
QY 122 agtgcaccgcgcacgc 139  
Db 2166 ACGAGCAGCCCGCGCAGC 2183

RESULT 5  
US-08-125-468-1  
Sequence 1, Application US/08125468  
Patent No. 5585385  
GENERAL INFORMATION:  
APPLICANT: Ryan, Michael J.  
APPLICANT: Lotvin, Jason A.  
APPLICANT: Strathy, Nancy  
APPLICANT: Fantini, Susan E.  
TITLE OF INVENTION: Cloning of the biosynthetic pathway for  
TITLE OF INVENTION: chlortetracycline and tetracycline formation and cosmid  
NUMBER OF SEQUENCES: 1  
CORRESPONDENCE ADDRESS:  
ADDRESS: American Cyanamid Company  
CITY: Wayne  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07470  
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/125,468  
FILING DATE: 22-SEP-1993  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Tsevdos, Estelle J  
REGISTRATION NUMBER: 31,145  
REFERENCE/DOCKET NUMBER: 31,255-02  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (201)831-3241  
TELEFAX: (201)831-3305  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3001 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-125-468-1  
Query Match 27.1%; Score 38.2; DB 1; Length 3001;  
Best Local Similarity 54.7%; Pred. No. 0.2;  
Matches 76; Conservative 0; Mismatches 63; Indels 0; Gaps 0;  
QY 3 ccgcgagcccttgctcttcactaccgcgcggtgagcgccgtgacgtgtgtacc 62  
Db 14962 CGACGTGCTACCGCGCGCGGACCACTCTGTGCGCGCGGATGCGGACCGCGA 15021  
QY 63 gctgagccgctgctgagcgagcgcaaggcctcgacggttcgtgcgggcttcggtca 122  
Db 15022 GCTGATCATCCGCGCGCGGAGGATGATGATGCTTTCGCGCGCGGCGCGGTCT 15081  
QY 123 gctgcaccgcgcacgcag 141  
Db 15082 GCTGCTGCTCCCGCAACAG 15100  
RESULT 6  
US-08-474-933-1  
Sequence 1, Application US/08474933  
Patent No. 5866410  
GENERAL INFORMATION:  
APPLICANT: Ryan, Michael J.  
APPLICANT: Lotvin, Jason A.  
APPLICANT: Strathy, Nancy  
APPLICANT: Fantini, Susan E.  
TITLE OF INVENTION: Cloning of the biosynthetic pathway for  
TITLE OF INVENTION: chlortetracycline and tetracycline formation and cosmid  
NUMBER OF SEQUENCES: 1  
CORRESPONDENCE ADDRESS:  
ADDRESS: American Cyanamid Company  
CITY: Wayne  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07470  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/474,933  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/125,468

FILING DATE: 22-SEP-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Tsevdos, Estelle J  
REGISTRATION NUMBER: 31,145  
REFERENCE/DOCKET NUMBER: 31,255-02  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (201)831-3241  
TELEFAX: (201)831-3305  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3001 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-474-933-1

Query Match 27.1%; Score 38.2; DB 2; Length 30001;  
Best Local Similarity 54.7%; Pred. No. 0.2;  
Matches 76; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

QY 3 cgcagagccctggctcttcgactaccgcgcggtggcgccgtgacgctgtcgtaacc 62  
Db 14962 CGACGTCGCTACGCGCGCGGCGACACCTCTGGTCGCGCGCGGATGCCGACCGGA 15021

QY 63 gctgagcgcgtgcggaggagcgcaaggcctcgagcgggtccgtgcgggcttcggtca 122  
Db 15022 GCTGATCACCGCGGACGCGCGGAGGATCAGCGATTGCTTCGCGCGGCGCGCGGTCT 15081

QY 123 gctgacccgcgcagcag 141  
Db 15082 GCTGCTGCTCCCGGACAG 15100

## RESULT 7

US-09-036-987A-25  
Sequence 25, Application US/09036987A  
Patent No. 6143526  
GENERAL INFORMATION:  
APPLICANT: Baltz, Richard H.  
APPLICANT: Broughton, Mary C.  
APPLICANT: Crawford, Kathryn P.  
APPLICANT: Madduri, Krishnamurthy  
APPLICANT: Merlo, Donald J.  
APPLICANT: Treadway, Patti J.  
APPLICANT: Turner, Jan R.  
APPLICANT: Waldron, Clive  
TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide  
TITLE OF INVENTION: Production  
NUMBER OF SEQUENCES: 39  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Dow Agrosciences LLC Patent Department  
STREET: 9330 Zionsville Road  
CITY: Indianapolis  
STATE: Indiana  
COUNTRY: USA  
ZIP: 46268  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/036, 987A  
FILING DATE: 09-MAR-1998  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Stuart, Donald R.  
REGISTRATION NUMBER: 28,479  
REFERENCE/DOCKET NUMBER: 50,608  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (317)337-4816

TELEFAX: (317)337-4847  
INFORMATION FOR SEQ ID NO: 25:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2310 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 88..1077  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1165..1992  
US-09-036-987A-25

Query Match 27.0%; Score 38; DB 4; Length 2310;  
Best Local Similarity 57.6%; Pred. No. 0.23;  
Matches 68; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 1 gccacgagccctggctcttcgactaccgcgcggtggcgccgtgacgctgtcgtaac 60  
Db 1213 GTCGCCGACGCGTGGGTTCTTCGCGGAGACGCGGAGACGCGGAGTGGACCGGTG 1272

QY 61 ccgctgagccctggcgaggagcgcaaggcctcgagcgggtccggtgccggggttcg 118  
Db 1273 GTGATCAACCGCGCGGTACACGCGGTGTGACGCGCGGAGTCCGACCGGACCGCG 1330

## RESULT 8

US-08-804-227C-1  
Sequence 1, Application US/08804227C  
Patent No. 5876991  
GENERAL INFORMATION:  
APPLICANT: DeHoff, Bradley S.  
APPLICANT: Kuhstoss, Stuart A.  
APPLICANT: Rostek, Paul R., Jr.  
APPLICANT: Sutton, Kimberly L.  
TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: THOMAS G. PLANT 1501  
STREET: LILLY CORPORATE CENTER  
CITY: INDIANAPOLIS  
STATE: IN  
COUNTRY: USA  
ZIP: 46285  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: ASCII(DOS) Text only  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/804,227C  
FILING DATE: February 21, 1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Plant, Thomas, G.  
REGISTRATION NUMBER: 35,784  
REFERENCE/DOCKET NUMBER: X-8231  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 317-276-2459  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 43280 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 816..14234



```
FEATURE:
NAME/KEY: CDS
LOCATION: 14351..19945
FEATURE:
NAME/KEY: CDS
LOCATION: 20010..31199
FEATURE:
NAME/KEY: CDS
LOCATION: 31232..36067
FEATURE:
NAME/KEY: CDS
LOCATION: 36249..41774
US-08-804-227C-1

Query Match 24.3%; Score 34.2; DB 2; Length 43280;
Best Local Similarity 55.5%; Pred. No. 1.7;
Matches 66; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

QY 1 qcgcgagccctgggtcctctgactaccgcggtggtggcgccgtgacgctgtgtac 60
DB 30450 GACGCGAGGGCTGCCCTGTCGACGCGCACAGGCGGCGGACGCGCTGCTGTG 30509

QY 61 ccgctgagcgcgctggcgagcgcgaagcctcgagcgggtccggtcccggttcgg 119
DB 30510 CCGATGGGCTCAACCGAGCGCCCTTCGCGCCTCGGAGCGATCACCCCTTCCTCAG 30568

RESULT 9
US-07-702-771-2/c
; Sequence 2, Application US/07702771
; Patent No. 5610276
; GENERAL INFORMATION:
; APPLICANT: Wong, Gail L.
; APPLICANT: McCormick, Frank
; TITLE OF INVENTION: Cloning and Purification of P62
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cetus Corporation
; STREET: 1400 Fifty-third Street
; CITY: Emeryville
; STATE: California
; ZIP: 94608-2997
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07702,771
; FILING DATE: 17-MAY-1991
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Glotta, Gregory J.
; REGISTRATION NUMBER: 32,028
; REFERENCE/DOCKET NUMBER: 2594
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-420-3300
; TELEFAX: 415-658-5239
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1485 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; HYPOTHETICAL: N
; ANTI-SENSE: N
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
US-07-702-771-2

FEATURE:
NAME/KEY: CDS
LOCATION: 14351..19945
FEATURE:
NAME/KEY: CDS
LOCATION: 20010..31199
FEATURE:
NAME/KEY: CDS
LOCATION: 31232..36067
FEATURE:
NAME/KEY: CDS
LOCATION: 36249..41774
US-08-804-227C-1

Query Match 24.1%; Score 34; DB 1; Length 1485;
Best Local Similarity 53.8%; Pred. No. 2;
Matches 70; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 5 ccgagccctgggtcctctgactaccgcggtggtggcgccgtgacgctgtgtaccgc 64
DB 143 CCGAGGCTGTGGCGAGGGGCGAGCAGCGGGGTGCGGCTGGCCGCCACTGTGCGGT 84

QY 65 tgagcgcgctggcgagcgcgaagcctcgacgagcgggtccggtccggttcggtcagc 124
DB 83 CGGACCGCGTGGCGGCGGCGAGCAGCGGTGGCGGTGCGTGGCGGCGAGGCCGCGG 24

QY 125 tgaccgcgcg 134
DB 23 CGCCCCCGCG 14

RESULT 10
US-08-438-883-2/c
; Sequence 2, Application US/0843883
; Patent No. 5731427
; GENERAL INFORMATION:
; APPLICANT: Wong, G.
; APPLICANT: McCormick, F.
; TITLE OF INVENTION: Purification and Cloning of p62
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/438,883
; FILING DATE: 10-MAY-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/702,771
; FILING DATE: 17-MAY-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Pochopien, Donald J.
; REGISTRATION NUMBER: 32,167
; REFERENCE/DOCKET NUMBER: 27527/33275
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6300
; TELEFAX: 312/474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1485 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; HYPOTHETICAL: N
; ANTI-SENSE: N
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
US-08-438-883-2

Query Match 24.1%; Score 34; DB 1; Length 1485;
Best Local Similarity 53.8%; Pred. No. 2;
Matches 70; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 5 ccgagccctgggtcctctgactaccgcggtggtggcgccgtgacgctgtgtaccgc 64
DB 143 CCGAGGCTGTGGCGAGGGGCGAGCAGCGGGGTGCGGCTGGCCGCCACTGTGCGGT 84
```

	Query Match	24.1%;	Score 34;	DB 1;	Length 1485;
	Best Local Similarity	53.8%;	Pred. No. 2;		
	Matches 70;	Conservative 0;	Mismatches 60;	Indels 0;	Gaps 0;
QY	5	ccgaggccctgggctctctgactaccgcgcgtggcgccgtgacgtgtcgtaccgc	64		
Db	143	CGAGGCTGTGGCCGAGGGGGACACCGGGTCTGGCGCTGGCCGCCACTGTGGCT	84		
QY	65	tgagcgcgcgtgggagagagcaaggcctcgagcgggtccgtgcgcggcttcggttcagc	124		
Db	83	CGGAGCCCGTGGCCGAGGCGCGCACACGCGTGGCGGCTGCCTGGCGGGCGAGCCCGGG	24		
QY	125	tgcaccgcgcg	134		

```
/ NAME: Kanski, Antoinette F.
/ REGISTRATION NUMBER: 34,202
/ REFERENCE/DOCKET NUMBER: 20344-20553.10
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (415) 813-5600
/ TELEFAX: (415) 494-0792
/ TELEX: 706141 MRSNFOERS SFO
/ INFORMATION FOR SEQ ID NO: 5:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 1212 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: CDNA
/ HYPOTHETICAL: NO
/ ANTI-SENSE: NO
/ ORIGINAL SOURCE:
/ ORGANISM: Homo Sapiens
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: 211..1212
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: 52..54
/ OTHER INFORMATION: /note= "Upstream in frame stop"
/ OTHER INFORMATION: codon
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: 98..119
/ OTHER INFORMATION: /note= "Oligonucleotide used for
/ OTHER INFORMATION: primer extension"
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (270-271)
/ OTHER INFORMATION: /note= "Position of the first
/ OTHER INFORMATION: intron and alternate sequences(SEQ ID NO:6 through SEQ ID NO:
/ OTHER INFORMATION: diverge"
/ US-08-449-933-5

Query Match 24.0%; Score 33.8; DB 2; Length 1212;
Best Local Similarity 58.4%; Pred. No. 2.2;
Matches 59; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

QY 8 aggcctgggtctcttctgactaccgcgcgtggcggtgacgtgtcgtagccgtga 67
Db 131 AGCCTCCGCTCCCGCGCTCTTCCGCGCCAGCGCGCGCCACCCCTCCCTCCGCGC 190
QY 68 gcgcctgctggggaggagcgaaggcctcgagcggttcgtg 108
Db 191 CCCCCGCGCGGGGAGGACATGCGCGGCACAGGCGGNG 231

RESULT 14
US-08-449-933-1
/ Sequence 1, Application US/08449933
/ Patent No. 5859195
/ GENERAL INFORMATION:
/ APPLICANT: Collins, Francis S.
/ APPLICANT: Wallace, Margaret R.
/ APPLICANT: Marchuk, Douglas A.
/ APPLICANT: Anderson, Lone B.
/ APPLICANT: Guttman, David H.
/ TITLE OF INVENTION: Neurofibromatosis Gene
/ NUMBER OF SEQUENCES: 14
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Morrison & Foerster
/ STREET: 755 Page Mill Road
/ CITY: Palo Alto
/ STATE: CA
/ COUNTRY: USA
/ ZIP: 94304-1018
/ COMPUTER READABLE FORM:

/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC Compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/449,933
/ FILING DATE: 25-MAY-1995
/ CLASSIFICATION: 435
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Kanski, Antoinette F.
/ REGISTRATION NUMBER: 34,202
/ REFERENCE/DOCKET NUMBER: 20344-20553.10
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (415) 813-5600
/ TELEFAX: (415) 494-0792
/ TELEX: 706141 MRSNFOERS SFO
/ INFORMATION FOR SEQ ID NO: 1:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 8937 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: CDNA to mRNA
/ HYPOTHETICAL: NO
/ ANTI-SENSE: NO
/ ORIGINAL SOURCE:
/ ORGANISM: Homo Sapiens
/ POSITION IN GENOME:
/ CHROMOSOME/SEGMENT: 17q11.2
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: 6814..8937
/ OTHER INFORMATION: /note= "Entire length of sequence
/ OTHER INFORMATION: clone p5"
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: 190..8646
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: 8425..8646
/ OTHER INFORMATION: /note= "219 nt PstI-HindIII
/ OTHER INFORMATION: fragment designated pMAL.B3A"
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: 382..1302
/ OTHER INFORMATION: /note= "918 nt HpaI-PstI fragment
/ OTHER INFORMATION: designated pMAL.HF3A.P"
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: 382..3909
/ OTHER INFORMATION: /note= "3523 nt HpaI-XhoI fragment
/ OTHER INFORMATION: designated pMAL.HF3A.X"
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: 8383..8937
/ OTHER INFORMATION: /note= "Clone B3A"
/ PUBLICATION INFORMATION:
/ AUTHORS: Wallace, M.R. et al.
/ TITLE: Type 1 Neurofibromatosis Gene: Correction
/ JOURNAL: Science
/ VOLUME: 250
/ ISSUE: 12/21/90
/ PAGES: 1749-
/ DATE: 12/21-1990
/ RELEVANT RESIDUES IN SEQ ID NO: 1: FROM 1 TO 8937
/ PUBLICATION INFORMATION:
/ AUTHORS: Wallace, M.R. et al.
/ TITLE: Type 1 Neurofibromatosis Gene: Identification
/ TITLE: of a Large Transcript in Three NFI Patients
/ JOURNAL: Science
/ VOLUME: 249
/ ISSUE: 07/13/90
/ PAGES: 181-186
```



Mon Jun 4 11:31:06 2001

us-09-331-723a-4.rni

Page 9



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 2, 2001, 00:00:32 ; Search time 1116.81 Seconds  
(without alignments)  
1861.953 Million cell updates/sec

Title: US-09-331-723A-4

Perfect score: 141

Sequence: 1 gccgcgagccctgggtc.....agctgacccgcgcacgcag 141

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1283235 seqs, 7373929652 residues

Total number of hits satisfying chosen parameters: 2566470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:\*

1: gb\_ba1:\*

2: gb\_ba2:\*

3: gb\_ba3:\*

4: gb\_in1:\*

5: gb\_in2:\*

6: gb\_in3:\*

7: gb\_om:\*

8: gb\_ov:\*

9: gb\_pat1:\*

10: gb\_pat2:\*

11: gb\_ph:\*

12: gb\_pil:\*

13: gb\_pl2:\*

14: gb\_pl3:\*

15: gb\_pl4:\*

16: em\_bal:\*

17: em\_ba2:\*

18: em\_fun:\*

19: em\_htgo\_hum:\*

20: em\_htgo\_inv:\*

21: em\_htgo\_rtd:\*

22: em\_htg\_hum1:\*

23: em\_htg\_hum2:\*

24: em\_htg\_hum3:\*

25: em\_htg\_hum4:\*

26: em\_htg\_hum5:\*

27: em\_htg\_hum6:\*

28: em\_htg\_hum7:\*

29: em\_htg\_hum8:\*

30: em\_htg\_inv1:\*

31: em\_htg\_inv2:\*

32: em\_htg\_inv3:\*

33: em\_htg\_rtd:\*

34: em\_hum1:\*

35: em\_hum2:\*

36: em\_hum3:\*

37: em\_hum4:\*

38: em\_hum5:\*

39: em\_hum6:\*

40: em\_hum7:\*

41: em\_in:\*

42: em\_om:\*

43: em\_or:\*

44: em\_ov:\*

45: em\_pat:\*

46: em\_ph:\*

47: em\_pi:\*

48: em\_ro:\*

49: em\_sts:\*

50: em\_sy:\*

51: em\_un:\*

52: em\_vi:\*

53: gb\_sts1:\*

54: gb\_sts2:\*

55: gb\_sts3:\*

56: gb\_sy:\*

57: gb\_un:\*

58: gb\_vil:\*

59: gb\_vil2:\*

60: gb\_vil3:\*

61: gb\_vil4:\*

62: gb\_vil5:\*

63: gb\_vil6:\*

64: gb\_vil7:\*

65: gb\_vil8:\*

66: gb\_vil9:\*

67: gb\_vil10:\*

68: gb\_vil11:\*

69: gb\_vil12:\*

70: gb\_vil13:\*

71: gb\_vil14:\*

72: gb\_vil15:\*

73: gb\_vil16:\*

74: gb\_vil17:\*

75: gb\_vil18:\*

76: gb\_vil19:\*

77: gb\_vil20:\*

78: gb\_vil21:\*

79: gb\_vil22:\*

80: gb\_vil23:\*

81: gb\_vil24:\*

82: gb\_vil25:\*

83: gb\_vil26:\*

84: gb\_vil27:\*

85: gb\_vil28:\*

86: gb\_vil29:\*

87: gb\_vil30:\*

88: gb\_vil31:\*

89: gb\_vil32:\*

90: gb\_vil33:\*

91: gb\_vil34:\*

92: gb\_vil35:\*

93: gb\_vil36:\*

94: gb\_vil37:\*

95: gb\_vil38:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	141	100.0	141	9	A78950
2	141	100.0	2480	12	AF068635
3	139.4	98.9	2573	9	A78956
4	139.4	98.9	3381	9	A78965
5	139.4	98.9	10265	12	AF030179
c 6	136.2	96.6	3383	9	A78933
c 7	39.4	27.9	37750	3	SC6D7
c 8	39.4	27.9	151089	13	AP003105
c 9	39	27.7	4767	1	AB021222
10	39	27.7	11403	1	AE004488
11	39	27.7	16903	1	AE004724

```

12 38.8 27.5 2900 3 PGLIP
13 38.8 27.5 2900 9 A16323
14 38.8 27.5 2900 9 AR038049
15 38.8 27.5 2900 10 I49947
16 38.4 27.2 7459 3 SCBA2
17 38.2 27.1 4422 3 STWCTS
18 38.2 27.1 30001 9 AR031172
19 38.2 27.1 30001 10 I32939
20 37 26.2 4674 2 AL047026
21 37 26.2 146921 13 AP002836
22 37 26.2 158133 13 AP002816
23 36.8 26.1 3777 1 AF139661
24 36.8 26.1 40744 3 SCJ4
25 36.6 26.0 9413 2 LDMGFA
26 36.6 26.0 10506 1 AE004806
27 36.6 26.0 11979 3 MXCRTNODA
28 36.6 26.0 40476 3 SCC75A
29 36.6 26.0 63554 14 H0410C08
30 36.4 25.8 12172 1 AE001872
31 36.4 25.8 13431 1 AE005044
32 36.4 25.8 33517 3 SC10B7
33 36.4 25.8 34158 3 SC2K31
34 36.4 25.8 40007 3 SC4B10
35 36.4 25.8 43632 3 SC6A5
36 36.4 25.8 281026 76 AC079572
37 36.2 25.7 938 6 LMB19KRP
38 36.2 25.7 32941 3 SCM10
39 36 25.5 1026 9 A70058
40 36 25.5 3217 3 SGSECA
41 36 25.5 6231 3 SCC117
42 36 25.5 8168 2 D87924
43 36 25.5 12198 1 AE001953
44 36 25.5 26500 3 SC9H11
45 36 25.5 36224 3 SCD78

```

## ALIGNMENTS

```

RESULT 1
A78950 141 bp DNA PAT 20-OCT-1999
LOCUS Sequence 4 from Patent WO9829554.
DEFINITION A78950
ACCESSION A78950
VERSION A78950.1 GI:6092078
KEYWORDS
SOURCE Chlamydomonas reinhardtii.
ORGANISM Chlamydomonas reinhardtii.
REFERENCE Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
AUTHORS Ishige,F. and Sato,R.
TITLE METHODS OF CONFERRING PPO-INHIBITING HERBICIDE RESISTANCE TO PLANTS
BY GENE MANIPULATION
JOURNAL Patent: WO 9829554-A 4 09-JUL-1998;
ISHIGE FUMIHARU (JP); SATO RYO (JP)
FEATURES
    source
        1..141
            /organism="Chlamydomonas reinhardtii"
            /strain="CC-407"
            /db_xref="taxon:3055"
            /protein_id="AAC79685.1"
            /protein_id="AAC79685.1"
            /db_xref="GI:3928794"
            /translation="MMLTQPTGATSSRRSIOIRSAAHVSAKAVAPRTPFVSAPATA
            ASPATAARRTLHRTAAATGAPTASGACVAKTLDNVYDVIIVGGSLGLVGTQALAA
            QHRIQNLVTEAREVGNVGIISMSGCVWEEGPNRPNDSMLQIADVDSCKDLVF
            GPTAPRFVWEGKLRVPPSGDAFTDLSIPKIRAGLGAIGLNGAMPSEESVE
            QFTRNLGDEVFFRLIEFFCSGVYAGDPSKLSKAAFRNIWLEKNGSLVGGAIKLF
            QERQSNAPPDRPLPKPKGTQVSGFRKGLKMLPDATERNIPDKIRVNNKIVSLGRE
            ADUGYGLVYDTPGRVRFVAFVAVLAPSVVADLVKQAPAAALGSDYPPVGVG
            TLYSLPSAVREERKASDGSVPFGQLHPTQTGTTTGLTYSLSLFFPGRAPEGHMLLN
            YIGGTTNRCINOTTEQLVQVDKDLRNVIKPDAPKPRVGVWVRAIPQPNLGLH
            EQLDKARKALDARGQLGVHLLGNVSGVALGKVEHGEYSANLAKSVSKAAVKA"
BASE COUNT 15 a 53 c 54 g 19 t
ORIGIN

```

```

Query Match 100.0%; Score 141; DB 9; Length 141;
Best Local Similarity 100.0%; Pred. No. 2.1e-14;
Matches 141; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggcgcgagccctggctccttcgactaccgcgcggtggcgccgtgacgtgtcgatc 60
DB 1 GCCTGAGCGCCCTGGCTCCTTCTGCTACCTCCGCGGCGCTGACGCTCTGCTAC 60

```

```

QY 61 ccgctgagccctggcgagagcgcaagcctcgagcggtcccggtcccggttcgggt 120
DB 61 CCGCTGAGCGCCCTGGCGAGGAGCGCAAGCGCTCGGACGGTCCCGGGGCTTCGGT 120
QY 121 cagctgcacccgcgcgcagcag 141
DB 121 CAGCTGACCCCGCGCAGCAG 141
RESULT 2
AF068635 2480 bp mRNA PLN 27-NOV-1998
LOCUS Chlamydomonas reinhardtii protoporphyrinogen oxidase precursor
DEFINITION (Ppx1) mRNA, nuclear gene encoding chloroplast protein, complete.
CDs
ACCESSION AF068635
VERSION AF068635.1 GI:3928793
KEYWORDS Chlamydomonas reinhardtii.
SOURCE Chlamydomonas reinhardtii
ORGANISM Chlamydomonadaceae; Chlamydomonas.
REFERENCE 1 (bases 1 to 2480)
AUTHORS Randolph-Anderson,B.L., Sato,R., Johnson,A.M., Harris,E.H.,
Hauser,C.R., Oeda,K., Ishige,F., Nishio,S., Gillham,N.W. and
Boynton,J.B.
TITLE Isolation and characterization of a mutant protoporphyrinogen
oxidase gene from Chlamydomonas reinhardtii conferring resistance
to porphyrin herbicides
JOURNAL Plant Mol. Biol. 38 (5), 839-859 (1998)
MEDLINE 99077317
REFERENCE 2 (bases 1 to 2480)
AUTHORS Randolph-Anderson,B.L., Sato,R., Johnson,A.M., Harris,E.H.,
Hauser,C.R., Oeda,K., Ishige,F., Nishio,S., Gillham,N.W. and
Boynton,J.B.
TITLE Direct Submission
JOURNAL Submitted (27-MAY-1998) DCMB Group, LSRC Building, Duke University,
Research Drive, Durham, NC 27708, USA
FEATURES
    source
        1..2480
            /organism="Chlamydomonas reinhardtii"
            /strain="CC-621"
            /db_xref="taxon:3055"
            /map="13.7 map units from srl, 12.3 map units from pfl6"
            /clone_lib="Lambda Zap II library; Wafenschmidt et al
            (1993) Plant Cell 5:809-820"
            /chromosome="linkage group IX"
            /note="wildtype strain"
            1..2480
            /gene="Ppx1"
            /gene="Ppx1"
            /function="prophyrin biosynthesis"
            /codon_start=1
            /product="protoporphyrinogen oxidase precursor"
            /protein_id="AAC79685.1"
            /db_xref="GI:3928794"
            /translation="MMLTQPTGATSSRRSIOIRSAAHVSAKAVAPRTPFVSAPATA
            ASPATAARRTLHRTAAATGAPTASGACVAKTLDNVYDVIIVGGSLGLVGTQALAA
            QHRIQNLVTEAREVGNVGIISMSGCVWEEGPNRPNDSMLQIADVDSCKDLVF
            GPTAPRFVWEGKLRVPPSGDAFTDLSIPKIRAGLGAIGLNGAMPSEESVE
            QFTRNLGDEVFFRLIEFFCSGVYAGDPSKLSKAAFRNIWLEKNGSLVGGAIKLF
            QERQSNAPPDRPLPKPKGTQVSGFRKGLKMLPDATERNIPDKIRVNNKIVSLGRE
            ADUGYGLVYDTPGRVRFVAFVAVLAPSVVADLVKQAPAAALGSDYPPVGVG
            TLYSLPSAVREERKASDGSVPFGQLHPTQTGTTTGLTYSLSLFFPGRAPEGHMLLN
            YIGGTTNRCINOTTEQLVQVDKDLRNVIKPDAPKPRVGVWVRAIPQPNLGLH
            EQLDKARKALDARGQLGVHLLGNVSGVALGKVEHGEYSANLAKSVSKAAVKA"
BASE COUNT 462 a 758 c 834 g 426 t
ORIGIN

```

```

Query Match 100.0%; Score 141; DB 12; Length 2480;
Best Local Similarity 100.0%; Pred. No. 4.6e-15;

```



Matches 141; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gccgcgagccctgggctcttccactaccgcgcgctgagccgctgacgtgctgtac 60  
|||||  
Db 1219 GCCGCCGAGGCCCTGGCTCTCTCGACTACCGCGCGGTGGCGCGCTGACGCTGTCGTAC 1278  
|||||  
QY 61 ccgctgagccgctgagcgagcgcgcaaggcctcgacgggtccgtgcccgggttcggt 120  
|||||  
Db 1279 CCGCTGAGCGCGGTGGCGGAGGAGCGCAAGGCCTCGGACGGGTCCTGCGCGGCTTCGGT 1338  
|||||

QY 121 cagctgcaccgcgcgcgcag 141  
|||||  
Db 1339 CAGCTGCACCGCGCGCAGCAG 1359  
|||||

RESULT 3  
A78956 2573 bp DNA PAT 20-OCT-1999  
LOCUS Sequence 10 from Patent WO9829554.  
DEFINITION A78956  
ACCESSION A78956  
VERSION A78956.1 GI:6092084  
KEYWORDS Chlamydomonas reinhardtii.  
SOURCE Chlamydomonas reinhardtii.  
ORGANISM Chlamydomonas reinhardtii.  
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;  
Chlamydomonadales; Chlamydomonas.  
REFERENCE 1 (bases 1 to 2573)  
AUTHORS Ishige, F. and Sato, R.  
TITLE METHODS OF CONFERRING PPO-INHIBITING HERBICIDE RESISTANCE TO PLANTS  
BY GENE MANIPULATION  
JOURNAL Patent: WO 9829554-A, 10 09-JUL-1998;  
ISHIGE FUMIHARU (JP); SATO RYO (JP)  
FEATURES  
source  
/organism="Chlamydomonas reinhardtii"  
/strain="RS-3"  
/db\_xref="taxon:3055"  
BASE COUNT 470 a 859 c 735 g 509 t  
ORIGIN

Query Match 98.9%; Score 139.4; DB 9; Length 2573;  
Best Local Similarity 99.3%; Pred. No. 8.3e-15;  
Matches 140; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 gccgcgagccctgggctcttccactaccgcgcgctgagccgctgacgtgctgtac 60  
|||||  
Db 255 GCCGCCGAGGCCCTGGCTCTCTCGACTACCGCGCGGTGGCGCGCTGACGCTGTCGTAC 314  
|||||  
QY 61 ccgctgagccgctgagcgagcgcgcaaggcctcgacgggtccgtgcccgggttcggt 120  
|||||  
Db 315 CCGCTGAGCGCGGTGGCGGAGGAGCGCAAGGCCTCGGACGGGTCCTGCGCGGCTTCGGT 374  
|||||

QY 121 cagctgcaccgcgcgcgcag 141  
|||||  
Db 375 CAGCTGCACCGCGCGCAGCAG 395  
|||||

RESULT 4  
A78965 3381 bp DNA PAT 20-OCT-1999  
LOCUS Sequence 19 from Patent WO9829554.  
DEFINITION A78965  
ACCESSION A78965  
VERSION A78965.1 GI:6092093  
KEYWORDS Chlamydomonas reinhardtii.  
SOURCE Chlamydomonas reinhardtii.  
ORGANISM Chlamydomonas reinhardtii.  
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;  
Chlamydomonadales; Chlamydomonas.  
REFERENCE 1 (bases 1 to 3381)  
AUTHORS Ishige, F. and Sato, R.  
TITLE METHODS OF CONFERRING PPO-INHIBITING HERBICIDE RESISTANCE TO PLANTS  
BY GENE MANIPULATION

JOURNAL Patent: WO 9829554-A, 10 09-JUL-1998;  
ISHIGE FUMIHARU (JP); SATO RYO (JP)  
FEATURES  
source  
Location/Qualifiers  
1..3381  
/organism="Chlamydomonas reinhardtii"  
/strain="RS-3"  
/db\_xref="taxon:3055"  
BASE COUNT 644 a 1058 c 1010 g 669 t  
ORIGIN

Query Match 98.9%; Score 139.4; DB 9; Length 3381;  
Best Local Similarity 99.3%; Pred. No. 7.2e-15;  
Matches 140; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 gccgcgagccctgggctcttccactaccgcgcgctgagccgctgacgtgctgtac 60  
|||||  
Db 255 GCCGCCGAGGCCCTGGCTCTCTCGACTACCGCGCGGTGGCGCGCTGACGCTGTCGTAC 314  
|||||  
QY 61 ccgctgagccgctgagcgagcgcgcaaggcctcgacgggtccgtgcccgggttcggt 120  
|||||  
Db 315 CCGCTGAGCGCGGTGGCGGAGGAGCGCAAGGCCTCGGACGGGTCCTGCGCGGCTTCGGT 374  
|||||

QY 121 cagctgcaccgcgcgcgcag 141  
|||||  
Db 375 CAGCTGCACCGCGCGCAGCAG 395  
|||||

RESULT 5  
AF030179 10265 bp DNA PLN 26-NOV-1998  
LOCUS Chlamydomonas reinhardtii herbicide-resistant protoporphylinogen  
DEFINITION oxidase precursor (Ppx1) gene, nuclear gene encoding chloroplast  
protein, complete cds.  
ACCESSION AF030179  
VERSION AF030179.1 GI:3928568  
KEYWORDS  
SOURCE Chlamydomonas reinhardtii.  
ORGANISM Chlamydomonas reinhardtii.  
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;  
Chlamydomonadales; Chlamydomonas.  
REFERENCE 1 (bases 1 to 10265)  
AUTHORS Randolph-Anderson, B.L., Sato, R., Johnson, A.M., Harris, E.H.,  
Hauser, C.R., Oeda, K., Ishige, F., Nishio, S., Gillham, N.W. and  
Boynton, J.E.  
TITLE Isolation and Characterization of a Mutant Protoporphyrinogen  
Oxidase Gene from Chlamydomonas reinhardtii Conferring Resistance  
to Porphyrin Herbicides  
JOURNAL Plant Mol. Biol. 38/5, 839-858 (1998)  
REFERENCE 2 (bases 1 to 10265)  
AUTHORS Randolph-Anderson, B.L., Sato, R., Johnson, A.M., Harris, E.H.,  
Hauser, C.R., Oeda, K., Ishige, F., Nishio, S., Gillham, N.W. and  
Boynton, J.E.  
TITLE Direct Submission  
JOURNAL Submitted (15-OCT-1997) DCMB Group, LSRC Building, Duke University,  
Research Drive, Durham, NC 27708, USA  
FEATURES  
source  
Location/Qualifiers  
1..10265  
/organism="Chlamydomonas reinhardtii"  
/strain="GB-2674"  
/db\_xref="taxon:3055"  
/map="13.7 map units from srl; 12.3 map units from pfl6"  
/chromosome="linkage group IX"  
/note="rs3 mutant strain"  
Hind 10 fragment  
163..164  
/note="location of approximately 3.8kb deletion"  
join(<1717..2163,2366..2410,2983..3124,3393..3442,  
3800..3885,4153..4221,4391..4449,4801..4870,5159..5312,  
5709..5855,6113..6256,7241..7333,7622..7718,8103..8191)  
/gene="Ppx1"  
/product="herbicide-resistant protoporphylinogen oxidase  
precursor"

```
gene
  <1717..>8191
  /gene="Ppx1"
  join(1717..2163,2366..2410,2983..3124,3393..3442,
3800..3885,4153..4221,4391..4449,4801..4870,5159..5312,
5709..5855,6113..6256,7241..7333,7622..7718,8103..8191)
  /gene="Ppx1"
  /function="porphyrin biosynthesis"
  /codon_start=1
  /product="herbicide-resistant protoporphyrinogen oxidase
precursor"
  /protein_id="AAC79630.1"
  /db_xref="GI:3928569"
  /translation="MWLTQPTGTAASSRQSRSSQJRSAAHVSAAKVAAPRPTFVSAPATA
ASPTAAARTRTHRTAAATGAPTASGACVAKTLDNVYDVIVVGGSLGTCGALAA
QHKQNFVTEARVYRNGITSMGSDGIVYWEQSPNFQNDMSLQIADVSCCKDLVF
GDTPTAFVWGGKLRPVPSGLDFTDLMSIFPKIRAGLGAIGLINGAMSPFEESVF
QFTNRNGDEVFRLTPEFCGVSAGVAGPSKLSMKAAFNRIWLEKNGSGIVGGAKLF
QERQSNPAPRDPRLPRPKPGTQVTSFSGKGLKMLPDAIERNIPDKIRVNWKLVSIGRE
ADGRYGLVDTPEGRVFAVAVALTAPSVVADLVKEQAPAAEALGSDFYPPMGAV
TLSVPLSAVREERKASDGSVFGQLHPTQGITTLGTTYSSTLPPGAPGPHMLLIN
YIGTNRGIVNQTTEOLVEQVDKDLNRMVVKPDAPKPRVGVVRVWRAIPQFNIGHL
EQDKARKALDAAGLQVHLGNGVSGVALGVKVEHGESAAANLAKSVSKAAVKA"
transit_peptide 1717..1962
  /gene="Ppx1"
  join(1963..2163,2366..2410,2983..3124,3393..3442,
3800..3885,4153..4221,4391..4449,4801..4870,5159..5312,
5709..5855,6113..6256,7241..7333,7622..7718,8103..8188)
  /gene="Ppx1"
  /product="herbicide-resistant protoporphyrinogen oxidase"
  /note="Ppx1"
  /note="encodes methionine rather than wildtype valine"
  /replace="g"
variation
  5751
  /gene="Ppx1"
  /note="encodes methionine rather than wildtype valine"
  /replace="g"
BASE COUNT 1903 a 3057 c 3413 g 1892 t
ORIGIN

Query Match 98.9%; Score 139.4; DB 12; Length 10265;
Best Local Similarity 99.3%; Pred. No. 4e-15;
Matches 140; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 gccgcgagccctggctcttcctgactaccgcgcggtggcgccgtgacgctgtcgtac 60
|||||
Db 5715 GCCGCCGAGCCCTGGGCTCTTCGACTACCGCGCGATGGCGCGGTGTCGTAC 5774
|||||

Qy 61 ccgctgagcgcgtgaggagagcgaagcctcgagcgcgtcgctggcggttcggt 120
|||||
Db 5775 CCCTGAGCGCGTGGGAGGAGCGCAAGCGCTCGGACGGTCCGTGCGCGCTTCGT 5834
|||||

Qy 121 cagctgaccgcgcgcagcag 141
|||||
Db 5835 CAGCTGCACCGCGCAGCAG 5855

RESULT 6
A79833/c
LOCUS A79833 3383 bp DNA PAT 20-OCT-1999
DEFINITION Sequence 1 from Patent WO9704088.
ACCESSION A79833
VERSION A79833.1 GI:6092737
KEYWORDS Chlamydomonas reinhardtii.
SOURCE Chlamydomonas reinhardtii.
ORGANISM Chlamydomonas reinhardtii.
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
Chlamydomonadales; Chlamydomonas.
1 (bases 1 to 3383)
Sato,R. and Boynton,J.
TITLE PORPHYRIN-ACCUMULATING TYPE HERBICIDE RESISTANCE GENE
JOURNAL Patent: WO 9704088-A 1 06-FEB-1997;
SUMITOMO CHEMICAL CO (JP); UNIV DUKE (US)
FEATURES
Location/Qualifiers
1..3383
/organism="Chlamydomonas reinhardtii"
```

```
/strain="RS-3"
/db_xref="taxon:3055"
BASE COUNT 671 a 1008 c 1057 g 647 t
ORIGIN

Query Match 96.6%; Score 136.2; DB 9; Length 3383;
Best Local Similarity 97.9%; Pred. No. 2.4e-14;
Matches 138; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 gccgcgagccctggctcttcctgactaccgcgcggtggcgccgtgacgctgtcgtac 60
|||||
Db 3128 GCCGCCGAGCCCTGGGCTCTTCGACTACCGCGCATGGCGCGGTGTCGTAC 3069
|||||

Qy 61 ccgctgagcgcgtgaggagagcgaagcctcgagcgcgtcgctggcggttcggt 120
|||||
Db 3068 CCGCTGAGCGCCCTGGGAGAGAGCGGACACCTCGGACGGTCCGTCGCCGGGTTCCGT 3009
|||||

Qy 121 cagctgaccgcgcgcagcag 141
|||||
Db 3008 CATCTGCACCGCGCAGCAG 2988
|||||

RESULT 7
SC6D7/c
LOCUS SC6D7 37750 bp DNA BCT 02-MAY-2000
DEFINITION Streptomyces coelicolor cosmid 6D7.
ACCESSION AL133213
VERSION AL133213.2 GI:7688318
KEYWORDS acd; acyl-CoA dehydrogenase (EC 1.3.99.-); ankyrin-like protein;
asnC-family transcriptional regulator; ATP
phosphoribosyltransferase; cbxX/CfGx family protein; chID;
chitinase; glycosyl transferase; GntR-family regulatory protein;
hisG; integral membrane protein; membrane protein;
phosphodiesterase; secreted protein; retK-family transcriptional
regulator; transmembrane transport protein.
SOURCE Streptomyces coelicolor A3(2).
ORGANISM Streptomyces coelicolor A3(2).
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
1 (bases 1 to 37750)
Redenbach,M., Kieser,H.M., Denapalte,D., Eichner,A., Cullum,J.,
Kinashi,H. and Hopwood,D.A.
A set of ordered cosmids and a detailed genetic and physical map
for the 8 Mb Streptomyces coelicolor A3(2) chromosome
Mol. Microbiol. 21 (1), 77-96 (1996)
97000351
2 (bases 1 to 37750)
Seeger,K.J. and Harris,D.
Unpublished
3 (bases 1 to 37750)
Thomson,N.R., Parkhill,J., Barrell,B.G. and Rajandream,M.A.
Submitted (24-NOV-1999) Streptomyces coelicolor sequencing project,
Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof.
David A. Hopwood. [3] John Innes Centre, Norwich Research Park,
Colney, Norwich, Norfolk NR4 7UH, UK
On May 3, 2000 this sequence version replaced gi:6468696.
Notes:
Streptomyces coelicolor sequencing at The Sanger Centre is funded
by the BBSRC and Beowulf Genomics
Details of S. coelicolor sequencing at the Sanger Centre are
available on the World Wide Web.
(URL: http://www.sanger.ac.uk/Projects/S.coelicolor/)
CDS are numbered using the following system eg SC7B7.01c. SC (S.
coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary
strand).
The more significant matches with motifs in the PROSITE database
are also included but some of these may be fortuitous.
The length in codons is given for each CDS.
Usually the highest scoring match found by fasta -o is given for
CDS which show significant similarity to other CDS in the database.
```

The position of possible ribosome binding site sequences are given where these have been used to deduce the initiation codon. Gene prediction is based on positional base preference in codons using a specially developed Hidden Markov Model (Krogh et al., Nucleic Acids Research, 22(22):4768-4778(1994)) and the FramePlot program of Bibb et al., Gene 30:157-66(1984) as implemented at <http://www.nih.go.jp/junc/cgi-bin/frameplot.pl>. CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, or att) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.

IMPORTANT: this sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions.

Cosmid 607. Overlaps with cosmid 1A8A on the AseI-g genomic restriction fragment.

#### FEATURES

```

source
    1. .37750
       Location/Qualifiers
         /organism="Streptomyces coelicolor A3(2)"
         /strain="A3(2)"
         /db_xref="taxon:100226"
         /clone="cosmid 6D7"
         complement(1..4276)
         /gene="SC6D7.32"
         complement(<1..4276)
         /gene="SC6D7.32"
         /note="SC6D7.32, unknown, len: > 1425 aa. Weakly similar
to several including: Streptomyces coelicolor
TR:Q93A4(EMBL:AL035478) putative secreted protein
SC265.19 (1156 aa), fasta scores opt: 327 z-score: 314.0
E(); 4.1e-10 28.0% identity in 938 aa overlap and
Rhodococcus capsulatus TR:O68032(EMBL:AF010496) potential
exonuclease (1238 aa), fasta scores opt: 439 z-score:
421.6 E(); 4.1e-16 28.8% identity in 1154 aa overlap. The
predicted protein product of this CDS is alanine rich and
contains multiple probable coiled-coils between residues:
369 to 401 (33 residues), 406 to 433 (28 residues), 483 to
519 (37 residues), 563 to 604 (42 residues), 850 to 921 (72
residues), 926 to 982 (57 residues), 1025 to 1054 (30
residues), 1072 to 1106 (35 residues), 1124 to 1167 (44
residues), 1199 to 1232 (34 residues) and 1278 to 1309 (32
residues)."
/codon_start=1
/transl_table=1
/product="hypothetical protein SC6D7.32."
/protein_id="CAB61684.1"
/db_xref="GI:6468728"
/translacion="MYELSRVRLYSIGPAGARYADTVLDLRGVRVLPDPAPTAQEFF
EEVPGPRPAPAGVLFLENGGKSVLLKLFVSLPGHRTNLGASGVLRFLFLA
DCGHVALEWHQVOTCEVYGVKYSNMGROVQSDPKFAEAWYSFRPGGLTDLNLP
VAETAIVRPEVCSAGCAGRRMTKGPEDALTGAKAYPHLSVHWEETHDRIEHLGD
LCILDPELFYQREMANDEGEANGLFVYKSDSDFTDLLLRVDTDRDTGLADLVGGCG
NKLGRAGLIAERDFTAGVSDLLGRIVEMAEARTDARVHTAERTNLRANRLTARG
AQEVRGLDIAQVRYAAVYVTHAESADRSNLIAELAYRHASLNLGAEKAAAGK
RELAERTLSAWQAAEAVLRHRAADRVARVSAIQEAEFDAAPALAAKAAVALV
RLTHAAGAKAEHANEARSAAIQEAGDAHRDSTTAETAQARSAGHRLQRLTE
VEQTEAVRAGVLDDSDAPDAPRAALASDAEKTVAADWTALETAGRATERHAEA
AAESRELTAAARAADAATAERAHEARTATAELAEELGLAEARGPIOPR
FDADDPRAARTADGALSPEELDFADELRELDLDTVSSAEROLFOLRYAAADDSR
ILGDGGLLLPGPDLATVEFLGEGIPALPGWRYLAQAVDPDADHARVLAARPELV
ILGLITDPGSRAREVIGDAALPRSAVAVGTAAALLAPAPAGSGTGVLPVPPNP
AMHDEQAEERHARARAGEDEETRALAARGKRELARLAEQEAOKARRAADALAG
ARTAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAE
LAFLRERAGQVYRLRELADEGAEAARQAELRAAEADDERAAQRAADADARLAR
ALARSEIAGAPDVPPEAGPGGAEPKSLPDLREAVRAASQYVKEVYGVADLRAR
QAESDEARADLDRLSNKVRFAOLLSPDGSQROAAARAAVQLLETR
MSSAEOLGRGEAEERHAPEDGTHDPELLPDAEHAQSLLRTAETLTLTEA
LARSREHALLDHAARAEDAGGDETAAMLRLLDHAPAEQEDPEPPAGDLEA
RRAAEARRSLRGCAADLSAAEAARASOVLVRHANSRTYEQVTPAPQOIRLEUPA
ALPHEAAKAWADAFAPRLVLTDELAQLERNKDSIVDLRLGLVESALATLRSRQLRSL

```

```

gene
    CDS
        /note="SC6D7.31, possible membrane protein, len: 293 aa.
contains a possible membrane spanning hydrophobic
domain."
        /codon_start=1
        /transl_table=1
        /product="putative membrane protein."
        /protein_id="CAB61683.1"
        /db_xref="GI:6468727"
        /translacion="MTEYVDORSEPEAGAPATAVTPADARADARLARVAFGLQKLLPA
RQVEYADLLRRYRDPFFARLADAVAGGLIVLEYSPRAGMAVSRAEVSFVPMGSD
YARRAATQDGRFLUHLAHLAAVAMAYPRPEDLADGGYIGRVTVNGVDAFVQAQCRRL
ERAPQCGENTDPAHLAAMAIWRRSATGATKARLPGSTTIGVRAAGFLT
ESGFLQRTGDHGGTYETTARYQLQVRDMAGSAALTELLDLGVVYVTDGTATLLPAEE
TDDLELVADAGLPFHS"
        complement(5272..5276)
        complement(5352..5278)
        /gene="SC6D7.30"
        complement(5352..5278)
        /gene="SC6D7.30"
        complement(5352..5278)
        /note="SC6D7.30, possible membrane protein, len: 508 aa.
contains a possible membrane spanning hydrophobic domain
and a probable coiled-coil between residues 259..286 (28
residues)"
        /codon_start=1
        /transl_table=1
        /product="putative membrane protein."
        /protein_id="CAB61682.1"
        /db_xref="GI:6468726"
        /translacion="MTSPASTDSVPPOLSAGPRPAPPADDEGLARLRALACAPLH
DDAKANLAGEYSVGYMAEVALAAIDLVLTNMDFTGADHDQIVARLIPRVAQAAPH
TYAARVAVNVLGALDVTSAOIAAEVKLEVLINRGLADAQALAAEQARYTVO
XSETLRALDATRNRVAVDMLSAVPMIAEALHDVADRYRHNALITNRKARDTE
DAHKRRAELVDIVKCIIRRHQQLSRLEAGPLFRAEQDQRAFATPLTSLGLDYG
HLVAPVLPVESALRYTDAFFARGTGLKRPVSVRGDVLILITPPVREHLGAEMP
EPDLATPDDSRFSEQLAAAMELDPHIDAPRSLGALLADARADALPYLVALLAA
HAASPVPVATYRQGEKPLFAVDQDELDDPEFGGADLVGTALLDAAGMAADRTAA"
        complement(6987..7682)
        /gene="SC6D7.29"
        complement(6987..7682)
        /note="SC6D7.29, putative GntR-family transcriptional
regulator, len: 231 aa. Similar to many other regulators
e.g. Bacillus licheniformis SW:GNTR_BACLI(EMBL:D31631)
glucanate operon transcriptional repressor (P28 protein)
(243 aa), fasta scores opt: 183 z-score: 234.8 E();
1.1e-05 31.2% identity in 138 aa overlap and Chelatobacter
heintzii SW:NTRA_CHEHE(EMBL:U39411) NTA operon
transcriptional regulator (210 aa), fasta scores opt: 222
z-score: 284.0 E(); 1.9e-08 31.8% identity in 195 aa
overlap. Contains a PS00043 Bacterial regulatory proteins,
gntR family signature and a Pflam match to entry PF00392
gntR, Bacterial regulatory proteins, gntR family. Contains
a putative helix-turn-helix motif between residues 44..65
(+2.58 SD)."
        /codon_start=1
        /transl_table=1
        /product="putative GntR-family transcriptional regulator."
        /protein_id="CAB61681.1"
        /db_xref="GI:6468725"
        /translacion="MAEQLTGLADRRALLGRTSTAERVSILRSRIAGEYFPPTRLS
EDISGALGVSNRTLRSEFRLTRELHNLNMGVFRVLTIVEDVIDYRSLVSEC
AVVRGIPPPYPLDGLAAEAVLDRGRSGDKMGVTARIHFHRELVALAGSRTEDEL
MPSVFAELRLAHFHVVDPRRLHIEPYIARNAILSTLESEREAARLAVYLENSLKR
VEVILRRVGGEG"
        complement(7437..7613)
        /misc_feature

```















Mon Jun 4 11:31:04 2001

us-09-331-723a-4.rge

Page 12

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 1, 2001, 13:45:48 ; Search time 19.54 Seconds  
(without alignments)  
281.923 Million cell updates/sec

Title:

Perfect score: 242

Sequence: 1 AEAALGSFDPVPGAVTSLY.....ERKASDGSVPFGQLHPRQTQ 47

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL15.\*

1: sp-archaea.\*

2: sp-bacteria.\*

3: sp-fungi.\*

4: sp-human.\*

5: sp-invertebrate.\*

6: sp-mammal.\*

7: sp-mhc.\*

8: sp-organelle.\*

9: sp-phage.\*

10: sp-plant.\*

11: sp-rodent.\*

12: sp-unclassified.\*

13: sp-vertebrate.\*

14: sp-virus.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	242	100.0	563	10 Q92TA7	Q92ta7 chlamydomon
2	239	98.8	563	10 Q970P9	Q970p9 chlamydomon
3	164.5	68.0	548	10 Q9SBI4	Q9Sbi4 nicotiana t
4	164.5	68.0	548	10 Q9SAX9	Q9Sax9 nicotiana t
5	157.5	65.1	557	10 Q64384	Q64384 solanum tub
6	142.5	58.9	555	10 Q9SPL6	Q9Spl6 cichorium i
7	141.5	58.5	535	10 Q9M629	Q9M629 zea mays (m
8	134.5	55.6	562	10 Q9LRI8	Q9Lri8 spinacia ol
9	133.5	55.2	545	10 Q9ZS15	Q9Zsi5 mycobacteri
10	133.5	55.2	284	2 Q33239	Q33239 mycobacteri
11	57.5	23.8	602	5 Q9V180	Q9Vv80 drosophila
12	57.5	23.8	1317	11 Q35211	Q35211 mus musculus
13	56	23.1	350	10 Q9PWW1	Q9Pww1 lampetra re
14	56	23.1	425	3 Q12741	Q12741 blastoclad
15	56	23.1	621	10 Q9M010	Q9m010 arabidopsi
16	56	23.1	1093	2 Q9K6B3	Q9k6b3 bacillus ha
17	55.5	22.9	1539	2 P96271	P96271 mycobacteri
18	55	22.7	343	2 Q9K202	Q9k202 streptomyce
19	55	22.7	439	2 P76130	P76130 escherichia

## ALIGNMENTS

RESULT	ID	PRELIMINARY	PRT	563 AA
Q92TA7	AC	Q92TA7		
DT	01-MAY-1999	(TREMBLrel. 10, Created)		
DT	01-MAY-1999	(TREMBLrel. 10, Last sequence update)		
DT	01-NOV-1999	(TREMBLrel. 12, Last annotation update)		
DE	PROTOPORPHYRINOGEN OXIDASE PRECURSOR.			
GN	PPX1.			
OS	Chlamydomonas reinhardtii.			
OC	Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;			
OC	Chlamydomonadaceae; Chlamydomonas.			
OX	NCBI_TaxID=3055;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=CC-621;			
RX	MEDLINE=99077317; PubMed=9862501;			
RA	Randolph-Anderson B.L., Sato R., Johnson A.M., Harris E.H.,			
RA	Hauser C.R., Oeda K., Ishige F., Nishio S., Gillham N.W.,			
RA	Boynton J.E.;			
RT	"Isolation and characterization of a mutant protoporphyrinogen oxidase gene from Chlamydomonas reinhardtii conferring resistance to porphyrin herbicides."			
RL	Plant Mol. Biol. 38:839-858(1998).			
DR	EMBL; AF068635; AAC79685.1;			
DR	MENDEL; 39099; Chire;3076;39099.			
SO	SEQUENCE 563 AA; 59802 MW; FC5E8FFC934CAF2 CRC64;			
Query Match 100.0%; Score 242; DB 10; Length 563;				
Best Local Similarity 100.0%; Pred. No. 2.5e-23;				
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;				
Qy	1	AEAALGSFDPVPGAVTSLYPLSANVEERKASDGSVPFGQLHPRQTQ 47		
Db	377	AEAALGSFDPVPGAVTSLYPLSANVEERKASDGSVPFGQLHPRQTQ 423		
RESULT 2				
Q9T0P9	AC	Q9T0P9	PRT	563 AA.
ID	Q9T0P9			
AC	Q9T0P9			

DT 01-MAY-2000 (TEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TEMBLrel. 13, Last sequence update)  
 DE HERBICIDE-RESISTANT PROTOPORPHYRINOGEN OXIDASE PRECURSOR.  
 GN PEX1.  
 OS Chlamydomonas reinhardtii.  
 OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;  
 OC Chlamydomonadaceae; Chlamydomonas.  
 OX NCBI\_TaxID=3055;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=GB-2674;  
 RA Randolph-Anderson B.L., Sato R., Johnson A.M., Harris E.H.,  
 RA Hauser C.R., Oeda K., Ishige F., Nishio S., Gillham N.W.,  
 RA Boynton J.E.;  
 RA "Isolation and Characterization of a Mutant Protoporphyrinogen Oxidase  
 RT Gene from Chlamydomonas reinhardtii Conferring Resistance to Porphyrin  
 RT Herbicides".  
 RL Plant Mol. Biol. 38:839-858(1998).  
 DR EMBL; AF030179; AAC79630.1; -.  
 KW Transit peptide.  
 FT TRANSIT 1 82 POTENTIAL.  
 FT CHAIN 83 563 HERBICIDE-RESISTANT PROTOPORPHYRINOGEN  
 FT OXIDASE.  
 FT VARIANT 389 389 M -> V.  
 FT SEQUENCE 563 AA; 59834 MW; 074584EC935CA3AF CRC64;  
 SQ  
 Query Match 98.8%; Score 239; DB 10; Length 563;  
 Best Local Similarity 97.9%; Pred. No. 6.2e-23;  
 Matches 46; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 AAALGSFDYPPVGVATLSYPLSAVREERKASDGSVPGFGLHPRQTQ 47  
 DB 377 AAALGSFDYPPVGVATLSYPLSAVREERKASDGSVPGFGLHPRQTQ 423  
 RESULT 3  
 Q9SB14 PRELIMINARY; PRT; 548 AA.  
 AC Q9SB14;  
 DT 01-MAY-2000 (TEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TEMBLrel. 13, Last sequence update)  
 DE PROTOPORPHYRINOGEN OXIDASE PX-1 (EC 1.3.3.4).  
 OS Nicotiana tabacum (Common tobacco).  
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 OC Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I;  
 OC Solanales; Solanaceae; Nicotiana.  
 OX NCBI\_TaxID=4097;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. SR1.  
 RA Horikoshi M., Mametsuka K., Hirooka T.;  
 RT "The Molecular Basis of Photobleaching Herbicide Resistance in  
 RT Tobacco".  
 RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF044128; AAD02290.1; -.  
 KW Oxidoreductase.  
 SQ SEQUENCE 548 AA; 59241 MW; 00E1A106F299E4AF CRC64;  
 Query Match 68.0%; Score 164.5; DB 10; Length 548;  
 Best Local Similarity 68.1%; Pred. No. 3e-13;  
 Matches 32; Conservative 6; Mismatches 1; Indels 1; Gaps 1;  
 QY 1 AAALGSFDYPPVGVATLSYPLSAVREERKASDGSVPGFGLHPRQTQ 47  
 DB 364 AAALSNFYPPVGVATISYQEAIRDERLV-DGELKGFGLHPRQTQ 409  
 RESULT 4  
 Q9SA99 PRELIMINARY; PRT; 555 AA.  
 AC Q9SA99;  
 DT 01-MAY-2000 (TEMBLrel. 13, Created)

ID Q9SA99 PRELIMINARY; PRT; 548 AA.  
 AC Q9SA99;  
 DT 01-MAY-2000 (TEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TEMBLrel. 13, Last sequence update)  
 DE PLASTIDAL PROTOPORPHYRINOGEN OXIDASE.  
 GN NTPPOX1.  
 OS Nicotiana tabacum (Common tobacco).  
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 OC Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I;  
 OC Solanales; Solanaceae; Nicotiana.  
 OX NCBI\_TaxID=4097;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=SAMSUN NN;  
 RA Watanabe N., Che F., Iwano M., Takayama S., Nakano T., Yoshida S.,  
 RA Isogai A.;  
 RA "Molecular characterization of photomixotrophic cultured tobacco cells  
 RT resistant to protoporphyrinogen oxidase-inhibiting herbicides".  
 RL Plant Physiol. 118:751-758(1998).  
 DR EMBL; AB020501; BAA34713.1; -.  
 SQ SEQUENCE 548 AA; 59215 MW; AB51A513FD3CE7BC CRC64;  
 Query Match 68.0%; Score 164.5; DB 10; Length 548;  
 Best Local Similarity 68.1%; Pred. No. 3e-13;  
 Matches 32; Conservative 6; Mismatches 1; Indels 1; Gaps 1;  
 QY 1 AAALGSFDYPPVGVATLSYPLSAVREERKASDGSVPGFGLHPRQTQ 47  
 DB 364 AAALSNFYPPVGVATISYQEAIRDERLV-DGELKGFGLHPRQTQ 409  
 RESULT 5  
 Q64384 PRELIMINARY; PRT; 557 AA.  
 ID Q64384;  
 AC Q64384;  
 DT 01-AUG-1998 (TEMBLrel. 07, Created)  
 DT 01-AUG-1998 (TEMBLrel. 07, Last sequence update)  
 DE PROTOPORPHYRINOGEN OXIDASE (EC 1.3.3.4).  
 OS Solanum tuberosum (Potato).  
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 OC Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I;  
 OC Solanales; Solanaceae; Solanum.  
 OX NCBI\_TaxID=4113;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. BINTJE;  
 RA Johnston D.J., Droz E., Roach J.D., Malnoe P.C.;  
 RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.  
 CC -1- CATALYTIC ACTIVITY: PROTOPORPHYRINOGEN-IX + O(2) = PROTOPORPHYRIN-  
 CC IX + H(2)O.  
 DR EMBL; AJ225107; CAAL2400.1; -.  
 DR MENDEL; 29307; Soltu; 3076; 29307.  
 KW Oxidoreductase.  
 SQ SEQUENCE 557 AA; 60463 MW; AE2B15B17E0B89F8 CRC64;  
 Query Match 65.1%; Score 157.5; DB 10; Length 557;  
 Best Local Similarity 66.0%; Pred. No. 2.5e-12;  
 Matches 31; Conservative 6; Mismatches 9; Indels 1; Gaps 1;  
 QY 1 AAALGSFDYPPVGVATLSYPLSAVREERKASDGSVPGFGLHPRQTQ 47  
 DB 373 AAALSNFYPPVAAVTISYQEAIRDERLV-DGELKGFGLHPRSQ 418  
 RESULT 6  
 Q9SPL6 PRELIMINARY; PRT; 555 AA.  
 ID Q9SPL6  
 AC Q9SPL6;  
 DT 01-MAY-2000 (TEMBLrel. 13, Created)



RX MEDLINE-20196006; PubMed-10731132:  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.R., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blaise R.G., Champe M., Pfaffler B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Farragkocch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brockstein P., Brotter P.,  
 RA Burtis K.C., Busam D.A., Butler H., Carden E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA DePaulos B., Delcher A., Deng Z., Mays A.D., Dew I., Datz S.M.,

QY 6 GSFDPVPGAVTLSYPLSAVR - EERKASDGSVPG 38

QY 6 GSFDPVPGAVTLSYPLSAVR - EERKASDGSVPG 38

```

DR PFAM: PF00433; pkinase_C: 1.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE: PS00108; PROTEIN_KINASE_ST; 1.
DR PROSITE: PS00111; PROTEIN_KINASE_DOM; 1.
DR SEQUENCE 425 AA; 47035 MW; 23A3A4D538768948 CRC64;

Query Match 23.1%; Score 56; DB 3; Length 425;
Best Local Similarity 33.9%; Pred. No. 31;
Matches 19; Conservative 4; Mismatches 17; Indels 16; Gaps 2;

QY 1 AAEALGSDYPPGVAVTLSYPLSAVREERK-----ASDGSVPFGQLH 43
DB 79 ATAHTNSDYSPPAATFSAFLDAVAERRRKTTLADLELRQLTGTGS---FGRVH 131

RESULT 15
Q9M010 PRELIMINARY; PRT; 621 AA.
ID Q9M010;
DC Q9M010;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE HYPOTHETICAL 68.6 KDA PROTEIN.
GN F7A7_180.
OS Arabidopsis thaliana (Mouse-ear cross).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Bevan M., Pohl T., Weizengger T., Bancroft I., Mewes H.W., Rudd S.,
RA Lemcke K., Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL161946; CAB82282.1; -.
KW Hypothetical protein.
SQ SEQUENCE 621 AA; 68579 MW; 259E20AAF74CE73F CRC64;

Query Match 23.1%; Score 56; DB 10; Length 621;
Best Local Similarity 38.9%; Pred. No. 47;
Matches 14; Conservative 7; Mismatches 13; Indels 2; Gaps 1;

QY 12 PVGAVTLSYPLSAVREERKASDGSVPFGQLHPRQT 47
DB 154 PVGSLEVS--LSSGGSDDSSAAASHPGFSENHPDVQ 187

Search completed: June 1, 2001, 13:47:36
Job time: 108 sec

```

Mon Jun 4 11:31:03 2001

us-09-331-723a-1.rspt

Page 6



GenCore version 4.5

Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 1, 2001, 13:45:58 ; Search time 8.09 Seconds  
(without alignments)  
199.012 Million cell updates/sec

Title: US-09-331-723A-1

Perfect score: 242

Sequence: 1 AALGSGFDYPPVGAATLSY.....ERKASDGVFCGLHPRTQ 47

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_39.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	161.5	66.7	548	1 PPOC_TOBAC	O24163 nicotiana t
2	133.5	55.2	537	1 PPOC_ARATH	P5826 arabidopsis
3	60.5	25.0	403	1 GCPE_SYNT3	P73672 synechocyst
4	60.5	25.0	690	1 LIP_STAOU	O0335 staphylococ
5	58	24.0	460	1 COAA_BPFI1	O80297 bacterioph
6	57	23.6	983	1 EPA3_CHICK	P29318 gallus gall
7	57	23.6	983	1 EPA3_HUMAN	P29320 homo sapien
8	57	23.6	984	1 EPA3_RAT	O08680 rattus norv
9	56	23.1	363	1 R23A_HUMAN	P54725 homo sapien
10	55.5	22.9	450	1 YPEB_BACSU	P38490 bacillus su
11	55.5	22.9	1434	1 NOS1_HUMAN	P29475 homo sapien
12	55	22.7	163	1 YAPF_SCHPO	O09791 schizosacch
13	55	22.7	363	1 R23A_MOUSE	P54726 mus musculu
14	54	22.3	3770	1 ACVS_EMEHI	P27742 emericella
15	53	21.9	288	1 PORB_METHH	O27771 methanobact
16	53	21.9	983	1 EPA3_MOUSE	P29319 mus musculu
17	53	21.9	987	1 EPB2_COTJA	Q90344 coturnix co
18	53	21.9	993	1 EPB2_MOUSE	P54763 mus musculu
19	53	21.9	1004	1 EPB2_CHICK	P28693 gallus gall
20	53	21.9	1055	1 EPB2_HUMAN	P29323 homo sapien
21	53	21.9	1569	1 GLI3_XENLA	Q10660 xenopus lae
22	52.5	21.7	942	1 YD73_SCHPO	O10328 schizosacch
23	52	21.5	234	1 CFI_VITVI	P51117 vitis vinif
24	52	21.5	323	1 JUND_CHICK	P27921 gallus gall
25	52	21.5	563	1 MDL1_PRUSE	P52706 prunus sero
26	52	21.5	787	1 SYFB_NEIMA	Q91va0 neisseria m
27	52	21.5	877	1 SYFB_MOUSE	Q60629 mus musculu
28	52	21.5	1005	1 EPA5_RAT	P54757 rattus norv
29	52	21.5	1013	1 EPA5_CHICK	P54755 gallus gall
30	52	21.5	1037	1 EPA5_HUMAN	P54756 homo sapien
31	52	21.5	3011	1 POLG_HCVH	P27958 h genome po
32	51.5	21.3	569	1 Y4FN_RHISN	P55452 rhizobium s
33	51	21.1	387	1 SOX1_HUMAN	O00570 homo sapien

34	50.5	20.9	450	1 LAT_NOCLA	Q05174 nocardia la
35	50.5	20.9	1586	1 SN22_HUMAN	P51531 homo sapien
36	50	20.7	344	1 F16P_SYN7	Q59943 synechococ
37	50	20.7	391	1 SOX1_MOUSE	P53783 mus musculu
38	50	20.7	605	1 WIS1_SCHPO	P33886 schizosacch
39	50	20.7	649	1 V70K_EPMV	P20129 eggplant mo
40	50	20.7	981	1 EPA3_BRARE	O13146 brachydanio
41	50	20.7	1002	1 EPB5_CHICK	Q07497 gallus gall
42	50	20.7	1106	1 ITA7_RAT	Q63258 rattus norv
43	49.5	20.5	229	1 CMF3_BACSU	P39147 bacillus su
44	49.5	20.5	341	1 JUND_MOUSE	P15066 mus musculu
45	49.5	20.5	341	1 JUND_RAT	P52909 rattus norv

## ALIGNMENTS

RESULT 1  
PPOC\_TOBAC  
ID PPOC\_TOBAC STANDARD; PRT: 548 AA.  
AC O24163;  
DT 15-DEC-1998 (Rel. 37, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE PROTOPORPHYRINOGEN OXIDASE, CHLOROPLAST PRECURSOR (EC 1.3.3.4) (PPO I)  
DE (PROTOPORPHYRINOGEN IX OXIDASE ISOZYME I) (PPX I).  
GN PPXI.  
OS Nicotiana tabacum (Common tobacco).  
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
OC Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I;  
OC Solanales; Solanaceae; Nicotiana.  
OX NCBI\_TaxID=4097;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=CV. SR1;  
RX MEDLINE=97385200; PubMed=9238074;  
RT Lermontova I., Kruse E., Mock H.-P., Grimm B.;  
RT "Cloning and characterization of a plastid and a mitochondrial  
isoform of tobacco protoporphyrinogen IX oxidase";  
RT Proc. Natl. Acad. Sci. U.S.A. 94:8895-8900(1997).  
CC -!- FUNCTION: CATALYZES THE 6-ELECTRON OXIDATION OF PROTOPORPHYRINOGEN  
IX TO FORM PROTOPORPHYRIN IX.  
CC -!- CATALYTIC ACTIVITY: PROTOPORPHYRINOGEN-IX + O(2) = PROTOPORPHYRIN-  
IX + H(2)O(2).  
CC -!- COFACTOR: (FAD) (BY SIMILARITY).  
CC -!- PATHWAY: PENULTIMATE STEP IN HEME AND PORPHYRIN BIOSYNTHESIS, AND  
CHLOROPHYLL SYNTHESIS.  
CC -!- SUBUNIT: HOMODIMER; CONTAINS ONE FAD PER HOMODIMER (BY  
SIMILARITY).  
CC -!- SUBCELLULAR LOCATION: CHLOROPLAST.  
CC -!- DEVELOPMENTAL STAGE: EXPRESSED IN EXPANDING PREMATURE LEAVES.  
CC -!- DECREASED EXPRESSION IN OLDEST LEAVES. EXPRESSED AT VERY LOW LEVEL  
IN ROOTS.  
CC -!- INDUCTION: OSCILLATING EXPRESSION DURING DIURNAL GROWTH. MAXIMAL  
EXPRESSION IN THE DARK PERIOD.  
CC -!- SIMILARITY: BELONGS TO THE PROTOPORPHYRINOGEN OXIDASE FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (see http://www.isb-sib.ch/announce/  
or send an email to license@isb-sib.ch).  
EMBL; Y13465; CAA73865.1;  
DR Protoporphyrinogen synthase; Heme biosynthesis; Oxidoreductase;  
KW Transit peptide; Flavoprotein; FAD; Chloroplast;  
KW Chlorophyll biosynthesis.  
FT TRANSIT 1 50 CHLOROPLAST (POTENTIAL).  
FT CHAIN 51 548 PROTOPORPHYRINOGEN OXIDASE.  
FT NP\_BIND 78 83 FAD (POTENTIAL).  
SEQUENCE 548 AA; 59230 MW; 66892E78F8A3E30 CRC64;



CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; D90908; BAA1717.1; -.  
 SQ SEQUENCE 403 AA; 44176 MW; 23948D183237DDB8 CRC64;

Query Match 25.0%; Score 60.5; DB 1; Length 403;  
 Best Local Similarity 43.2%; Pred. No. 1.2;  
 Matches 16; Conservative 6; Mismatches 10; Indels 5; Gaps 2;

QY 12 PVGAVTL--SYPL---SAVREERKASDGVPGFGQLH 43  
 DB 27 PVGAVTGGHVVQVQSMINEDLVDGSGVAGIRRLH 63

RESULT 4  
 ID LIP\_STAUA STANDARD; PRT; 690 AA.  
 AC P10335;  
 DT 01-MAR-1989 (Rel. 10, Created)  
 DT 01-MAR-1989 (Rel. 10, Last sequence update)  
 DT 01-OCT-1994 (Rel. 30, Last annotation update)  
 DE LIPASE PRECURSOR (EC 3.1.1.3) (GLYCEROL ESTER HYDROLASE).  
 GN GER.

OS Staphylococcus aureus.  
 CC Bacteria; Firmicutes; Bacillus/Clostridium group;  
 CC Bacillus/Staphylococcus group; staphylococcus.  
 OX NCBI\_TaxID=1280;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=86195821; PubMed=3009394;  
 RA Lee C.Y., Iandolo J.J.;  
 RT "Lysogenic conversion of staphylococcal lipase is caused by insertion  
 RT of the bacteriophage L54a genome into the lipase structural gene";  
 RL J. Bacteriol. 156:385-391(1986).  
 RN [2]  
 RP PROCESSING, AND SEQUENCE OF 296-307.

KC STRAIN=105;  
 RX MEDLINE=92193269; PubMed=1548232;  
 RA Roloff J., Normark S.;  
 RT "In vivo processing of Staphylococcus aureus lipase.";  
 RL J. Bacteriol. 174:1844-1847(1992).  
 CC -!- CATALYTIC ACTIVITY: TRIACYLGLYCEROL + H(2)O = DIACYLGLYCEROL +  
 CC A FATTY ACID ANION.  
 CC -!- SUBCELLULAR LOCATION: SECRETED.  
 CC -!- MISCELLANEOUS: THE EXPRESSION OF STAPHYLOCOCCUS LIPASE IS  
 CC NEGATIVELY REGULATED BY BACTERIOPHAGE LYSOGENIZATION (LIPASE  
 CC CONVERSION). PARTIAL WITH OTHER LIPASES (PANCREATIC, GASTRIC,  
 CC HEPATIC, LINGUAL, LIPOPROTEIN, BACTERIAL, ETC.).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; M12715; AAA26633.1; -.  
 DR PIR; A24545; A24545.  
 DR LaterPro; IPR000734; -.  
 DR PROSITE; PS00120; LIPASE\_SER; 1.  
 KW Hydrolase; Lipid degradation; Zymogen; Signal.  
 FT SIGNAL 1 37  
 FT PROPEP 38 295  
 FT CHAIN 296 690  
 ET LIPASE.

FT DOMAIN 311 690 HYDROPHOBIC.  
 FT ACT\_SITE 412 412 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 645 645 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 SQ SEQUENCE 690 AA; 76388 MW; 8E8E3654D0E01A3B CRC64;

Query Match 25.0%; Score 60.5; DB 1; Length 690;  
 Best Local Similarity 33.3%; Pred. No. 2.2;  
 Matches 15; Conservative 6; Mismatches 19; Indels 5; Gaps 1;

QY 8 FDYPPVGAVTLSTPLSAV-----REERKASDGVPGFGQLHPTQ 47  
 DB 573 YENPDLTGFFLMATTSRIIHDAREWRKNDGVVPISSLSLPSNQ 617

RESULT 5  
 ID COAA\_BP1F1 STANDARD; PRT; 460 AA.  
 AC O80297;  
 DT 13-JUL-1999 (Rel. 38, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE COAT PROTEIN A PRECURSOR (ATTACHMENT PROTEIN).  
 GN III OR 3.  
 OS Bacteriophage If1.  
 CC Viruses; ssDNA viruses; Inoviridae; Inovirus.  
 OX NCBI\_TaxID=10868;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Hill D.F., Hughes G., McNaughton J.C., Stockwell P.A., Petersen G.B.;  
 RT "DNA sequence of the filamentous coliphage If1";  
 RT Submitted (OCT-1993) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: COAT PROTEIN A IS NECESSARY FOR ADSORPTION OF THE VIRION  
 CC ONTO THE P-PILUS OF THE HOST CELL (BY SIMILARITY).  
 CC [1]

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; U02303; AAC62155.1; -.  
 KW Phage recognition; Coat protein; Signal.  
 FT SIGNAL 1 19 POTENTIAL.  
 FT CHAIN 20 460 COAT PROTEIN A.  
 SQ SEQUENCE 460 AA; 48790 MW; 9AD651968C715AB2 CRC64;

Query Match 24.0%; Score 58; DB 1; Length 460;  
 Best Local Similarity 43.6%; Pred. No. 3.1;  
 Matches 17; Conservative 6; Mismatches 10; Indels 6; Gaps 2;

QY 6 GSFDPPVG-----AVTLSTPLSAVREERKAS-DGSVPG 38  
 DB 305 GDFYVKMANAKDALTSPDLSALQADTGASLDGSGVG 343

RESULT 6  
 ID EPA3\_CHICK STANDARD; PRT; 983 AA.  
 AC P29318;  
 DT 01-DEC-1992 (Rel. 24, Created)  
 DT 01-DEC-1992 (Rel. 24, Last sequence update)  
 DE EPHRIN TYPE-A RECEPTOR 3 PRECURSOR (EC 2.7.1.112) (TYROSINE-PROTEIN  
 DE KINASE RECEPTOR ETK1) (CEK4).  
 GN EPA3 OR ETK1 OR CEK4.  
 OS Gallus gallus (Chicken).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 CC Gallus.



```

DR InterPro: IPR001090;
DR InterPro: IPR001245;
DR InterPro: IPR001426;
DR InterPro: IPR001660;
DR InterPro: IPR001777;
DR Pfam: PF01404; EPH1bd; 1.
DR Pfam: PF00536; SAM; 1.
DR Pfam: PF00041; fn3; 2.
DR Pfam: PF00069; pkinase; 1.
DR PRINTS: PR00014; FNTYPEII.
DR PRINTS: PR00109; TYRKINASE.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE: PS00111; PROTEIN_KINASE_DOM; 1.
DR PROSITE: PS00790; RECEPTOR_TYR_KIN_V_1; 1.
DR PROSITE: PS00791; RECEPTOR_TYR_KIN_V_2; 1.
DR PROSITE: PS01186; EGF_2; UNKNOWN_1.
KW Transferase: Tyrosine-protein kinase; ATP-binding; Phosphorylation;
KW Receptor; Transmembrane; Glycoprotein; Signal; Repeat.
FT SIGNAL 1 20
FT CHAIN 21 983
FT DOMAIN 21 541
FT TRANSMEM 542 565
FT DOMAIN 566 983
FT DOMAIN 189 322
FT DOMAIN 323 432
FT DOMAIN 433 530
FT DOMAIN 621 882
FT DOMAIN 909 983
FT SITE 981 983
FT NP_BIND 627 635
FT BINDING 633 653
FT ACT_SITE 746 746
FT MOD_RES 596 596
FT MOD_RES 602 602
FT MOD_RES 779 779
FT CARBOHYD 232 232
FT CARBOHYD 337 337
FT CARBOHYD 391 391
FT CARBOHYD 404 404
FT CARBOHYD 453 493
FT CONFLICT 507 507
FT CONFLICT 724 724
SQ SEQUENCE 983 AA; 110086 MW; B8D900FA80FF5121 CRC64;

Query Match 23.6%; Score 57; DB 1; Length 983;
Best Local Similarity 34.8%; Pred. No. 9.9;
Matches 16; Conservative 8; Mismatches 16; Indels 6; Gaps 2;

QY 2 AEALGSFDYPPV---GAVILSYPLSAVREERKASDGSVFGQLH 43
DB 671 ASIMGQDFHNIIRLEGVTKSPVMIVTEYME--NGSLDSFLRKH 714

RESULT 8
ID EPA3_RAT STANDARD; PRT; 984 AA.
AC O08680;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE EPHRIN TYPE-A RECEPTOR 3 (last annotation update)
DE KINASE RECEPTOR REK4 (TYRO-4).
GN EPA3 OR REK4 OR TYRO4.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OC NCBI_Taxid=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY;
RA MEDLINE=98120505; PubMed=9458884;

```

Li Y.Y., McTiernan C.F., Feldman A.M.;  
 "IL-1 beta alters the expression of the receptor tyrosine kinase gene  
 r-Epha3 in neonatal rat cardiomyocytes.";  
 Am. J. Physiol. 274:H331-H341(1998).  
 CC -1- FUNCTION: RECEPTOR FOR MEMBERS OF THE EPHRIN-A FAMILY. BINDS TO  
 EPHRIN-A2, -A3, -A4 AND -A5.  
 CC -1- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE - ADP +  
 PROTEIN TYROSINE PHOSPHATE.  
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
 CC -1- TISSUE SPECIFICITY: MOST ABUNDANT IN THE HEART, BRAIN, AND LUNG.  
 CC -1- INDUCTION: DOWNREGULATED BY INTERLEUKINE-1BETA IN NEONATAL CARDIAC  
 MYOCYTES.  
 CC -1- SIMILARITY: CONTAINS 1 SAM DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.  
 CC -1- SIMILARITY: BELONGS TO THE TYR FAMILY OF PROTEIN KINASES. EPHRIN  
 RECEPTOR SUBFAMILY.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: U69278; AAC06273.1; -  
 DR HSSP: P16109; 1FSB.  
 DR InterPro: IPR000561; -  
 DR InterPro: IPR000719; -  
 DR InterPro: IPR001090; -  
 DR InterPro: IPR001245; -  
 DR InterPro: IPR001426; -  
 DR InterPro: IPR001660; -  
 DR InterPro: IPR001777; -  
 DR InterPro: IPR01404; EPH1bd; 1.  
 DR Pfam: PF00536; SAM; 1.  
 DR Pfam: PF00041; fn3; 2.  
 DR Pfam: PF00069; pkinase; 1.  
 DR PRINTS: PR00014; FNTYPEII.  
 DR PRINTS: PR00109; TYRKINASE.  
 DR PROSITE: PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE: PS00109; PROTEIN\_KINASE\_TYR; 1.  
 DR PROSITE: PS00111; PROTEIN\_KINASE\_DOM; 1.  
 DR PROSITE: PS00790; RECEPTOR\_TYR\_KIN\_V\_1; 1.  
 DR PROSITE: PS00791; RECEPTOR\_TYR\_KIN\_V\_2; 1.  
 DR PROSITE: PS01186; EGF\_2; UNKNOWN\_1.  
 DR Transferase: Tyrosine-protein kinase; ATP-binding; Phosphorylation;  
 KW Receptor; Transmembrane; Glycoprotein; Signal; Repeat.  
 FT SIGNAL 1 20  
 FT CHAIN 21 984  
 FT DOMAIN 21 541  
 FT TRANSMEM 542 565  
 FT DOMAIN 566 984  
 FT DOMAIN 189 322  
 FT DOMAIN 323 431  
 FT DOMAIN 436 528  
 FT DOMAIN 622 883  
 FT DOMAIN 910 984  
 FT SITE 982 984  
 FT NP\_BIND 628 636  
 FT BINDING 634 654  
 FT ACT\_SITE 747 747  
 FT MOD\_RES 597 597  
 FT MOD\_RES 603 603  
 FT MOD\_RES 780 780  
 FT CARBOHYD 232 232  
 FT CARBOHYD 337 337  
 FT CARBOHYD 391 391  
 FT CARBOHYD 404 404  
 FT CARBOHYD 493 493  
 FT SEQUENCE 984 AA; 110227 MW; F170C49312F7A0AB CRC64;

Query Match 23.6%; Score 57; DB 1; Length 984;  
Best-Local Similarity 34.8%; Pred. No. 9.9;  
Matches 16; Conservative 8; Mismatches 16; Indels 6; Gaps 2;

QY 2 AEAALGSDYPPV-----GAVTLSPYLSAVREERKASDGSVFGQLH 43  
DB 572 ASIMQFDPHPIIRLEGGVTKSPVMIVTEYME--NGSLDSFLRXH 715

RESULT 9  
R23A\_HUMAN  
ID R23A\_HUMAN STANDARD; PRT; 363 AA.  
AC P54725;

DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE UV EXCISION REPAIR PROTEIN RAD23 HOMOLOG A (HHR23A).  
GN RAD23A.

OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;

RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=94222030; PubMed=8168482;  
RA Masutani C., Sugawara K., Yanagisawa J., Sonoyama T., Ui M.,  
RA Enomoto T., Takio K., Tanaka K., van der Spek P.J., Bootsma D.,  
RA Hoeljmakers J.H.J., Hanaoka F.;  
RT "Purification and cloning of a nucleotide excision repair complex  
involving the xeroderma pigmentosum group C protein and a human  
homologue of yeast RAD23";  
RL EMBO J. 13:1831-1843(1994).  
RN [2]

RN SEQUENCE FROM N.A.  
RP Lamerdin J., McCreedy P., Stillwagen S., Ramirez M., Carrano A.;  
RA Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.  
RN [3]

RN STRUCTURE BY NMR OF 319-363.  
RX MEDLINE=99061330; PubMed=9846873;  
RA Dieckmann T., Withers-Ward E.S., Jarosinski M.A., Liu C.F.,  
RA Chen I.S.Y., Feigon J.;  
RT "Structure of a human DNA repair protein UBA domain that interacts  
with HIV-1 Vpr";  
RL Nat. Struct. Biol. 5:1042-1047(1998).  
RN [4]

RN STRUCTURE BY NMR OF 319-363.  
RX MEDLINE=20541363; PubMed=11087358;  
RA Withers-Ward E.S., Mueller T.D., Chen I.S.Y., Feigon J.;  
RT "Biochemical and structural analysis of the interaction between the  
UBA(2) domain of the DNA repair protein HHR23A and HIV-1 Vpr";  
RL Biochemistry 39:14103-14112(2000).

CC 1-1 FUNCTION: COULD BE INVOLVED IN DNA EXCISION REPAIR.

CC 1-1 SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).

CC 1-1 SIMILARITY: CONTAINS 1 UBIQUITIN-LIKE DOMAIN.

CC 1-1 SIMILARITY: CONTAINS 2 UBA DOMAINS.

CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL: D21235; BAA04767.1; -

DR EMBL: AD000092; AAB51177.1; -

DR FDB: IDV0; 11-FEB-00.

DR MIM: 600061; -

DR InterPro: IPR000449; -

DR InterPro: IPR000626; -

DR Pfam: PF00627; UBA; 2.

DR Pfam: PF00240; ubiquitin; 1.

DR PROSITE: PS50053; UBIQUITIN\_2; 1.

KW DNA damage; DNA repair; Nuclear protein; 3D-structure.  
FT DOMAIN 1 81 UBIQUITIN-LIKE.  
SQ SEQUENCE 363 AA; 39609 MW; C4E47E9313BB47B5 CRC64;

Query Match 23.1%; Score 56; DB 1; Length 363;  
Best-Local Similarity 35.5%; Pred. No. 4.4;  
Matches 11; Conservative 6; Mismatches 14; Indels 0; Gaps 0;

QY 7 SFDYPPVGAVTLSYPLSAVREERKASDGSVP 37  
DB 100 STSFPAPTSGMSPHPPAREDKSPSEESAP 130

RESULT 10  
YPEB\_BACSU  
ID YPEB\_BACSU STANDARD; PRT; 450 AA.

AC P38490; P40774;

DT 01-OCT-1994 (Rel. 30, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 30-MAY-2000 (Rel. 39, Last annotation update)

DE HYPOTHETICAL 51.2 KDA PROTEIN IN RECO-CMK INTERGENIC REGION (PSPA12).

GN YPEB OR JOEB.

OS Bacillus subtilis.

OC Bacteria; Firmicutes; Bacillus/Clostridium group;

OC Bacillus/Staphylococcus group; Bacillus.

OX NCBI\_TaxID=1423;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=168 / MARBURG;

RX MEDLINE=96349105; PubMed=8760912;

RA Sorokin A.V., Azevedo V., Zumstein E., Galleron N., Ehrlich S.D.,

RA Serror P.;

RT "Sequence analysis of the Bacillus subtilis chromosome region between

the serA and kog loci cloned in a yeast artificial chromosome.";

RL Microbiology 142:2005-2016(1996).

RN [2]

RP SEQUENCE OF 220-450 FROM N.A.

RC STRAIN=168 / MARBURG;

RA Sorokin A.V., Azevedo V., Zumstein E., Galleron N., Ehrlich S.D.,

RA Serror P.;

RL Submitted (JUN-1994) to the EMBL/GenBank/DBJ databases.

RN [3]

RP SEQUENCE OF 1-331 FROM N.A.

RC STRAIN=168;

RX MEDLINE=96427343; PubMed=8830707;

RA Moriyama K., Hattori A., Miyata S., Kudoh S., Makino S.;

RT "A gene (slsB) encoding a spore cortex-lytic enzyme from Bacillus

subtilis and response of the enzyme to L-alanine-mediated

germination.";

RL J. Bacteriol. 178:6059-6063(1996).

RN [4]

RP SEQUENCE OF 1-27 FROM N.A.

RX MEDLINE=99108019; PubMed=3145906;

RA Smith H., de Jong A., Bron S., Venema G.;

RT "Characterization of signal-sequence-coding regions selected from the

Bacillus subtilis chromosome.";

RL Gene 70:351-361(1988).

CC 1-1 CAUTION: REF.4 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A

FRAMESHIFT IN POSITION 17.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL: L47648; AAC83958.1; -

DR EMBL: U11687; AAA85146.1; -

DR EMBL: D79978; BAA11474.1; -

DR EMBL: M22904; AAA22820.1; ALT\_FRAME.









Repeat: Phosphopantetheine.

FT	REPEAT	321	910	DOMAIN 1 (ADIPATE-ACTIVATING).
FT	REPEAT	1413	1593	DOMAIN 2 (CYSTEINE-ACTIVATING).
FT	REPEAT	2494	3078	DOMAIN 3 (VALINE-ACTIVATING).
FT	DOMAIN	850	919	ACYL CARRIER (ACP).
FT	DOMAIN	1929	2002	ACYL CARRIER (ACP).
FT	DOMAIN	3020	3087	ACYL CARRIER (ACP).
FT	BINDING	882	882	PHOSPHOPANTHETHEINE (BY SIMILARITY).
FT	BINDING	1965	1965	PHOSPHOPANTHETHEINE (BY SIMILARITY).
FT	BINDING	3050	3050	PHOSPHOPANTHETHEINE (BY SIMILARITY).
FT	ACT-SITE	3623	3623	THIOESTERASE (BY SIMILARITY).
FT	SEQUENCE	3770 AA;	42244 MW;	CB66B6D23A58CB0 CRC64;
QY		8	FDYPPGVAVTLPSLSAVREERKASDGSVPGQLH 43	
DD		3388	FEIPLNVETLSOGTASIRELCRVAPARGIGTGSLY 3423	
RESULT	15			
AD	PORB-METH			
IC	PORB-METH	STANDARD;	PRT;	288 AA.
AC	O27771;			
DT	30-MAY-2000	(Rel. 39, Created)		
DT	30-MAY-2000	(Rel. 39, Last sequence update)		
DT	30-MAY-2000	(Rel. 39, Last annotation update)		
DE	PYRUVATE SYNTHASE SUBUNIT PORB	(EC 1.2.7.1)		(PYRUVATE OXIDOREDUCTASE
DE	BETA CHAIN) (POR)	(PYRUVIC-FERREDOXIN OXIDOREDUCTASE BETA SUBUNIT).		
DE	PORB OR MTHL738.			
GN	Methanobacterium thermoautotrophicum.			
OS	Archaea; Euryarchaeota; Methanobacteriales; Methanobacteriaceae;			
OC	Methanothermobacter.			
CC	NCBI_TaxID=145262;			
OR	[1]			
RC	SEQUENCE FROM N.A.			
RC	STRAIN=DELTA H;			
RC	MEDLINE=98037514;	PubMed=9371463;		
RR	Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,			
RR	Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,			
RR	Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D.,			
RR	Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,			
RR	Jiwani N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,			
RR	McDougall S., Shimer G., Goyal A., Pietrovski S., Church G.M.,			
RR	Daniels C.J., Mao J.-I., Rice P., Nolling J., Reeve J.N.;			
RR	"Complete genome sequence of Methanobacterium thermoautotrophicum			
RR	deltaH: functional analysis and comparative genomics.";			
RR	J. Bacteriol. 179:7135-7155(1997).			
CCC	-1- CATALYTIC ACTIVITY: PYRUVATE + COA + OXIDIZED FERREDOXIN =			
CCC	ACETYL-COA + CO(2) + REDUCED FERREDOXIN.			
CCC	-1- SUBUNIT: HETEROTETRAMER OF ONE ALPHA, ONE BETA, ONE DELTA AND ONE			
CCC	GAMMA CHAIN.			
CCC	-----			
CCC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CCC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -			
CCC	the European Bioinformatics Institute. There are no restrictions on its			
CCC	use by non-profit institutions as long as its content is in no way			
CCC	modified and this statement is not removed. Usage by and for commercial			
CCC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announcement/">http://www.isb-sib.ch/</a>			
CCC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
CCC	-----			
CC	EMBL: AE000929; AAB86208.1; -			
KW	Oxidoreductase.			
SQ	SEQUENCE	286 AA;	31594 MW;	BB7DD458A62A21 CRC64;
Query Match		21.9%;	Score 53;	DB 1; Length 288;
Best Local Similarity		33.3%;	Pred. No. 8.6;	
Matches	11; Conservative	6; Mismatches	16; Indels	0; Gaps 0;

Mon Jun 4 11:31:03 2001

us-09-331-723a-1.rsp

Page 10

QY 11 PPVGAVTLSYPLSAVREERKASDGSVPGFGLH 43  
Db 173 PYVATASISYPEDFMEKVRKAKETDGPAYIHLH 205

Search completed: June 1, 2001, 13:47:48  
Job time: 110 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 1, 2001, 13:45:18 ; Search time 12.98 seconds  
(without alignments)  
248.843 Million cell updates/sec

Title:

Perfect score:

Sequence:

1 AAALGSGFDYPPVGAUTLSY.....ERKASDGSVPFGQLHPTQ 47

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 198801 seqs, 68722935 residues

Total number of hits satisfying chosen parameters: 198801

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR\_67.\*

1: pir2.\*

2: pir2.\*

3: pir3.\*

4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	161.5	66.7	548	2 T04058	protoporphyrinogen
2	157.5	65.1	557	2 T07116	protoporphyrinogen
3	133.5	55.2	545	2 T02005	protoporphyrinogen
4	60.5	25.0	403	2 S77159	gcpg protein - Syn
5	60.5	25.0	690	2 A24545	triacylglycerol li
6	57.5	23.8	284	2 E70506	hypothetical prote
7	57.5	23.8	1317	2 T03748	apoptosis associat
8	57	23.6	983	2 B45583	receptor tyrosine
9	57	23.6	983	2 A38224	protein-tyrosine k
10	56	23.1	363	2 S44443	RAD23 protein homo
11	56	23.1	425	2 S41099	protein kinase (EC
12	56	23.1	621	2 T48187	hypothetical prote
13	56	23.1	1093	2 H84126	cation efflux syst
14	55.5	22.9	450	2 E69934	conserved hypotet
15	55.5	22.9	1433	2 G01946	nitric-oxide synth
16	55.5	22.9	1539	2 G70630	probable ctpH prot
17	55	22.7	163	2 S62444	conserved hypotet
18	55	22.7	439	2 F64902	hypothetical prote
19	54.5	22.5	152	2 T51796	hypothetical prote
20	54	22.3	143	2 G72492	hypothetical prote
21	54	22.3	283	2 S75639	hypothetical prote
22	54	22.3	3770	2 A40889	delta-(L-alpha-ami
23	53	21.9	190	2 B81042	conserved hypotet
24	53	21.9	288	2 D69099	pyruvate synthase
25	53	21.9	303	2 S76535	probable phosphoe
26	53	21.9	478	2 A57174	protein-tyrosine k
27	53	21.9	480	2 I48760	protein-tyrosine k
28	53	21.9	970	2 I78842	receptor protein-t
29	53	21.9	983	2 A45583	receptor tyrosine

#### ALIGNMENTS

##### RESULT 1

T04058

protoporphyrinogen oxidase (EC 1.3.3.4) IX precursor, chloroplast - common tobacco  
C:Species: Nicotiana tabacum (common tobacco)  
C:Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 21-Jul-2000  
C:Accession: T04058

R:Lernontova, I.; Kruse, E.; Mock, H.P.; Grimm, B.  
Proc. Natl. Acad. Sci. U.S.A. 94, 8895-8900, 1997

A:Title: Cloning and characterization of a plastidial and a mitochondrial isoform of t  
A:Reference number: M15186; MUID:97385200  
A:Accession: T04058  
A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-548 <LDR>

A:Cross-references: EMBL:Y13465; NID:g2370332; PIDN:CAA73865.1; PID:g2370333  
A:Experimental source: strain SRI  
C:Genetics:  
A:Gene: ppxi

A:Function:  
A:Pathway: tetrapyrrole synthesis

A:Keywords: chloroplast; oxidoreductase

F:1-50/Domain: transit peptide (chloroplast) #status predicted <TNP>

F:51-548/Product: protoporphyrinogen oxidase IX #status predicted <MAT>

Query Match 66.7%; Score 161.5; DB 2; Length 548;  
Best Local Similarity 66.0%; Pred. No. 8.5e-14;  
Matches 31; Conservative 8; Mismatches 7; Indels 1; Gaps 1;

Qy 1 AAALGSGFDYPPVGAUTLSYPLSAVREERKASDGSVPFGQLHPTQ 47  
Db 364 AAALSNFYPPVGAUTITTPQAIKRLV-DGELKGFGLHPTQ 409

##### RESULT 2

T07116

protoporphyrinogen oxidase (EC 1.3.3.4) IX, chloroplast - potato  
C:Species: Solanum tuberosum (potato)  
C:Date: 30-Apr-1999 #sequence\_revision 30-Apr-1999 #text\_change 08-Oct-1999  
C:Accession: T07116

R:Johnston, D.J.

submitted to the EMBL Data Library, April 1998

A:Reference number: Z15932

A:Accession: T07116

A:Status: translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-557 <JOH>

A:Cross-references: EMBL:AJ225107; NID:e1289157; PIDN:CAA12400.1; PID:e1289158

A:Experimental source: cv. Bintje

C:Genetics:

A:Genome: nuclear

C:Function:

A:Pathway: tetrapyrrole biosynthesis  
C:Keywords: chloroplast; oxidoreductase

Query Match 65.1%; Score 157.5; DB 2; Length 557;  
Best Local Similarity 66.0%; Pred. No. 3e-13;  
Matches 31; Conservative 6; Mismatches 9; Indels 1; Gaps 1;

QY 1 AAALGSDYPPVAVTLVPLSAVREERKASDGVPGQLHPRQTQ 47  
DB 373 ADALSSFFYPVAVTISYPPQEAIRDERLV-DGELKGFGLHPRSQ 418

## RESULT 3

T02005  
Protoporphyrinogen oxidase (EC 1.3.3.4) - Arabidopsis thaliana  
N:Alternate names: protein T15B16.13  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 26-Feb-1999 #sequence\_revision 26-Feb-1999 #text\_change 21-Jul-2000  
C:Accession: T02005; JC5488

R:Stonking, T.; Smith, R.  
Submitted to the EMBL Data Library, November 1998  
A:Description: The sequence of A. thaliana T15B16.  
A:Reference number: Z14488

A:Accession: T02005  
A>Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-545 <STO>

A:CROSS-references: EMBL:AF104919; NID:g3859590; PID:g3859604

A:Experimental source: cultivar Columbia

A:Narita, S.; Tanaka, R.; Ito, T.; Okada, K.; Taketani, S.; Inokuchi, H.

Gene 182, 169-175, 1996

A:Title: Molecular cloning and characterization of a cDNA that encodes protoporphyrinogen  
A:Reference number: JC5488; MUID:97136707  
A:Accession: JC5488

A:Molecule type: mRNA

A:Residues: 1-217,226-545 <NAR>

A:CROSS-references: DDBJ:D83139; NID:g1183006; PIDN:BAA11820.1; PID:g18777018

C:Genetics:

A:Map position: 4

A:Introns: 126/3; 188/3; 225/1; 274/2; 301/2; 358/3; 412/1; 446/3

A>Note: T15B16.13

C:Function:

A:Description: catalyzes the removal of six electrons from protoporphyrinogen to generate

A:Pathway: chlorophyll biosynthesis; heme biosynthesis

C:Keywords: oxidoreductase

Query Match 55.2%; Score 133.5; DB 2; Length 545;  
Best Local Similarity 59.6%; Pred. No. 5e-10;  
Matches 28; Conservative 4; Mismatches 14; Indels 1; Gaps 1;

QY 1 AAALGSDYPPVAVTLVPLSAVREERKASDGVPGQLHPRQTQ 47  
DB 361 AAALSKLYPPVAVTISYPPQEAIRTE-CLIDGELKGFGLHPRQTQ 406

## RESULT 4

S77159  
gcpE protein - Synechocystis sp. (strain PCC 6803)  
N:Alternate names: protein slr2136  
C:Species: Synechocystis sp.

A:Variety: PCC 6803

C:Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 20-Jun-2000

C:Accession: S77159

R:Kaneko, T.; Sato, K.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;

O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda

DNA Res. 3, 109-136, 1996

A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis

s.

A:Reference number: S74322; MUID:97061201

A:Accession: S77159

A>Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-403 <KAN>

A:CROSS-references: EMBL:D90908; GB:AB001339; NID:g1652725; PIDN:BAA17717.1; PID:g165

A>Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

C:Genetics:

A:Gene: gcpE

C:Superfamily: gcpE protein

Query Match 25.0%; Score 60.5; DB 2; Length 403;  
Best Local Similarity 43.2%; Pred. No. 2.4; Indels 5; Gaps 2;  
Matches 16; Conservative 6; Mismatches 10;

QY 12 PVGAVTL-SYPL---SAVREERKASDGVPGQLH 43  
DB 27 PVGAVTGGGHPVVQSMINEDTLDVDSVAGIRRLH 63

## RESULT 5

A24545

triacylglycerol lipase (EC 3.1.1.3) - Staphylococcus aureus

C:Species: Staphylococcus aureus

C:Date: 30-Jun-1988 #sequence\_revision 30-Jun-1988 #text\_change 08-Oct-1999

C:Accession: A24545

R:Lee, C.Y.; Iandolo, J.J.

J. Bacteriol. 166, 385-391, 1986

A:Title: Lysogenic conversion of staphylococcal lipase is caused by insertion of the

A:Reference number: A24545; MUID:86195821

A:Molecule type: DNA

A:Residues: 1-690 <LEE>

A:CROSS-references: GB:M12715; NID:g153019; PIDN:AAA26633.1; PID:g153020

C:Superfamily: Staphylococcus triacylglycerol lipase

C:Keywords: carboxylic ester hydrolase

Query Match 25.0%; Score 60.5; DB 2; Length 690;  
Best Local Similarity 33.3%; Pred. No. 4.5;  
Matches 15; Conservative 6; Mismatches 19; Indels 5; Gaps 1;

QY 8 FDYPPVAVTLVPLSAV-----REERKASDGVPGQLHPRQTQ 47  
DB 573 YENDLGTFLMATLSRIIGHDAREWKNKNDGVVPPVLSLHPSNQ 617

## RESULT 6

E70506

hypothetical protein RV2734 - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis

C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 22-Oct-1999

C:Accession: E70506

R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon

; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holtroyd,

Nature 393, 537-544, 1998

A:Authors: Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno

A:Reference number: A70500; MUID:96295987

A:Accession: E70506

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-284 <COL>

A:CROSS-references: GB:Z98209; GB:AL123456; NID:g3261838; PIDN:CAB10910.1; PID:e33229

A:Experimental source: strain H37RV

C:Genetics:

A:Gene: RV2734

Query Match 23.8%; Score 57.5; DB 2; Length 284;  
Best Local Similarity 40.5%; Pred. No. 4.1;  
Matches 15; Conservative 8; Mismatches 9; Indels 5; Gaps 2;

QY 15 AVTLSPVLSAVREERKASDGVPGQLHPRQT 46

32 AMTLAKRLKAMGSDKYQTDGDPRTSGPGFVTHPRS 68  
Db

## RESULT 7

T03748  
apoptosis associated tyrosine kinase - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 24-Mar-1999 #sequence\_revision 24-Mar-1999 #text\_change 08-Oct-1999  
C:Accession: T03748  
R:Gaozza, E.; Baker, S.J.; Vora, R.K.; Reddy, E.P.  
Oncogene 15, 3127-3135, 1997  
A:Title: AATYK: A novel tyrosine kinase induced during growth arrest and apoptosis of my  
A:Reference number: Z15052; MUID:98105706  
A:Accession: T03748  
A:Status: preliminary; translated from GB/EMBL/DBDJ  
A:Molecule type: mRNA  
A:Residues: 1-1317 <GAO>  
A:Cross-references: EMBL:AF011308; NID:g2459992; PID:g2459993  
C:Genetics:  
A:Gene: AATYK

Query Match 23.8%; Score 57.5; DB 2; Length 1317;  
Best Local Similarity 50.0%; Pred. No. 24;  
Matches 17; Conservative 3; Mismatches 11; Indels 3; Gaps 2;

**QY** 6 GSDYPPVGA<sup>TLS</sup>PLSAVR-EERKASDGSVPG 38  
I : :||| | | | | | : | | | |

**Dd** 440 GAEEYPPSGAA--SSPGSAARLEICAPDSSPPG 471

## RESULT 8

receptor tyrosine kinase Cex4 - chicken  
C:Species: Gallus gallus (chicken)  
C:Date: 22-Apr-1993 #sequence\_revision 18-Nov-1994 #text\_change 18-Jun-1999  
C:Accession: B45583  
C:Author: R.Sajjadi, F.G.; Pasquale, E.B.; Subramani, S.  
New Biol. 3, 769-778, 1991  
A:Title: Identification of a new eph-related receptor tyrosine kinase gene from mouse an  
A:Reference number: A45583; MUID:92031278

Query Match 23.6%; Score 57; DB 2; Length 983;  
Best Local Similarity 34.8%; Pred. No. 20;  
Matches 16; Conservative 8; Mismatches 16; Indels 6; Gaps 2;

QY 2 AEALGSFDYPV---GAVTLSYPLSAVEREKASDGVPGQLH 43  
| : | | | : | | | | : | : | : |  
Db 671 ASIMGOEDHPNIIRLEGVTKSPVMIVTEYME--NGSLDSFLRKH 714

## RESULT

protein-tyrosine kinase (EC 2.7.1.112) hek precursor - human  
 C:Species: Homo sapiens (man)  
 C:Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 18-Jun-1999  
 C:Accession: A38224  
 R:Wicks, I.P.; Wilkinson, D.; Salvaris, E.; Boyd, A.W.  
 Proc. Natl. Acad. Sci. U.S.A. 89, 1611-1615, 1992  
 A:Title: Molecular cloning of HEK, the gene encoding a receptor tyrosine kinase expressed  
 A:Reference number: A38224: MUID:92179243

A:Accession: A38224  
A:Molecule type: mRNA  
A:Residues: 1-983 <NC>  
A:Cross-references: GB:M83941; NID:g183931; PIDN:AAAS86633.1.; PID:g183932  
A:Experimental source: pre-B-cell leukemia cell line LK63  
A:Note: sequence extracted from NCBI backbone (NCBIP:86627)

Query Match	23.6%	Score 57;	DB 2;	Length 983;
Best Local Similarity	34.8%	Pred. No. 20;		
Matches 16; Conservative	8;	Mismatches 16;	Indels 6;	Gaps 2;

QY 2 AEALGSFDYPPV---GAVTLSYPLSAVREERKASDGSVPFGQLH 43  
 } : | | | : | | | : | | : : | : | :  
 Db 671 ASIMGQFDHPNIIIRLEGVVTKSPVMIVTME--NGSLDSFLRKH 714

## RESULT 10

S4443  
 RAD23 protein homolog2 - human  
 N:Alternate names: HHR23A protein  
 C:Species: Homo sapiens (man)  
 C:Date: 25-Dec-1994 #sequence\_revision 01-Sep-1995 #text\_change 18-Feb-2000  
 R:Accession: S44443; R45076  
 R:Masutani, C.; Sugawara, K.; Yangisawa, J.; Sonoyama, T.; Ui, M.; Enomoto, T.; Taki  
 EMBO J. 13, 1931-1843, 1994  
 A:Title: Purification and cloning of a nucleotide excision repair complex involving t  
 A:Reference number: S44345; MUID:94222030  
 A:Accession: S44443  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-363 <MAS>  
 A:Cross-references: EMBL:D21235; NID:G498145; PID:BAA04767.1; PID:d1005299; PID:g498  
 R:Lamerding, J.; McCready, P.; Stillwagen, S.; Ramirez, M.; Carrano, A.  
 Submitted to the EMBL Data Library, November 1996  
 A:Description: Characterization by genomic sequence analysis of a gene-rich 111 kb re  
 A:Reference number: Z22306

Query Match 23.1%; Score 56; DB 2; Length 363;  
Best Local Similarity 35.5%; Pred. No. 8.7;  
Matches 11: Conservative 6; Mismatches 14; Indels 0; Gaps 0;

QY 7 SFDYPPVGAVTLSYPLSAVREERKASDGSVP 37  
| :|| :|| | ||:: | :|  
Db 100 STSFPPAPTSGMSHPPPAAREDKSPSEESAP 130

DEPT T 11

RESULT  
SA1099

protein kinase (EC 2.7.1.37). cAMP-dependent, catalytic chain C - fungus (Blasotocladia)

C:Species: Blasotocladia emersonii  
C:Date: 19-Mar-1997 #sequence\_revision 06-Jun-1997 #text\_change 28-May-1999  
C:Accession: S41099; S77889; S77890  
R:Francisco de Oliveira, J.C.; Cantisani Borges, A.C.; do Valle Marques, M.; Lopes Gomes, S.  
Eur. J. Biochem. 219, 555-562, 1994

A:Title: Cloning and characterization of the gene for the catalytic subunit of cAMP-depe

A:Reference number: S41099; MUID:94139736

A:Accession: S41099

A:Molecule type: DNA

A:Residues: 1-425 <FRA>

A:Cross-references: GB:L17008; MID:g304272; PIDN:AAA20074.1; PID:g304273

A:Accession: S77889

A:Molecule type: mRNA

A:Residues: 22-425 <FRB>

A:Cross-references: GB:M81709; GB:L17038; MID:g507140; PIDN:AAA19440.1; PID:g507141

A:Accession: S77890

A:Molecule type: protein

A:Residues: 2-16 <FRC>

C:Genetics:

A:Introns: 209/3; 243/3; 315/1

A:Superfamily: kinase-related transforming protein; protein kinase homology

C:Keywords: ATP; cAMP binding; magnesium; phosphoprotein; phosphotransferase; serine/thr

F:2-425/Product: protein kinase, cAMP-dependent, catalytic chain C #status experimental

F:114-370/Domain: protein kinase homology <KIN>

F:122-130/Region: protein kinase ATP-binding motif

F:127,128,194,200,243,256/Binding site: Mg-ATP (Phe, Gly, Glu, Glu, Thr) #status pr

F:145,164,239,241/Active site: Lys, Glu, Asp, Lys #status predicted

F:244,257/Binding site: magnesium (Asn, Asp) #status predicted

F:270/Binding site: phosphate (Thr) (covalent) #status predicted

Query Match 23.1%; Score 56; DB 2; Length 425;  
Best Local Similarity 33.9%; Pred. No. 10;  
Matches 19; Conservative 4; Mismatches 17; Indels 16; Gaps 2;

QY 1 AAALGSGFDYPPGVATLSYPLSAVREERK-----ASDGSVPFGQLH 43  
DB 79 ATAHTNSDYSPPAATPSAPLDAVAERRRKTTADLELRQTIGTGS---FGRVH 131

RESULT 12

T48187

hypothetical protein F7A7.180 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 20-Apr-2000

C:Accession: T48187

R:Bevan, M.; Terry, N.; Ardiles, W.; Buysschaert, C.; Dasseville, R.; De Clerck, R.; De

ewes, H.W.; Rudd, S.; Lemcke, K.; Mayer, K.F.X.

submitted to the Protein Sequence database, March 2000

A:Reference number: 224487

A:Accession: T48187

A:Molecule type: DNA

A:Residues: 1-621 <BEV>

A:Cross-references: EMBL:AL161946

A:Experimental source: cultivar Columbia; BAC clone F7A7

C:Genetics:

A:Map position: 5

A:Introns: 65/3; 263/3; 289/3; 403/1; 469/3; 495/2; 529/2; 592/3

A:Note: F7A7.180

Query Match 23.1%; Score 56; DB 2; Length 621;  
Best Local Similarity 38.9%; Pred. No. 16;  
Matches 14; Conservative 7; Mismatches 13; Indels 2; Gaps 1;

QY 12 PVGAVTSLYPLSAVREERKASDGSVPFGQLHPRTQ 47  
DB 154 PVGSLEVS--LSSGSDSSAAASHPGFSEHPDVQ 187

RESULT R3

H84126  
cation efflux system BH3816 [imported] - Bacillus halodurans (strain C-125)  
C:Species: Bacillus halodurans  
C:Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 31-Dec-2000  
C:Accession: H84126  
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; H

Nucleic Acids Res. 28, 4317-4331, 2000

A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans a

A:Reference number: A83650; MUID:20263314

A:Accession: H84126

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1093 <STO>

A:Cross-references: GB:AF001520; GB:BA000004; MID:g10176401; PIDN:BA007535.1; GSPDB:G

A:Experimental source: strain C-125

C:Genetics:

A:Gene: BH3816

Query Match 23.1%; Score 56; DB 2; Length 1093;  
Best Local Similarity 58.8%; Pred. No. 30;  
Matches 10; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 11 PPVGAVTSLYPLSAVRE 27  
DB 40 PPVGAVTASYPGASABE 56

RESULT 14

E69934

conserved hypothetical protein ypeB - Bacillus subtilis

C:Species: Bacillus subtilis

C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 01-Dec-2000

C:Accession: E69934; T44771

R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber

C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;

A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari,

Nature 390, 249-256, 1997

A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gal

iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hulio, M

Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino

A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau

Y.M.; Ogawa, K.; Ogihara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete

Rieser, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanl

A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Se

akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiya

T.; Wintere, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida

A:Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.

A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtili

A:Reference number: A69580; MUID:98044033

A:Accession: E69934

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-450 <KUN>

A:Cross-references: GB:Z99115; GB:Z99116; GB:AL009126; MID:g2634723; PIDN:CAB14224.1;

A:Experimental source: strain 168

R:Moriyama, R.; Hattori, A.; Miyata, S.; Kudoh, S.; Makino, S.

J. Bacteriol. 178, 6059-6063, 1996

A:Title: A gene (steB) encoding a spore cortex-lytic enzyme from Bacillus subtilis an

A:Reference number: Z22836; MUID:96427343

A:Accession: T44771

A>Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-331 <MOR>

A:Cross-references: EMBL:D79978; MID:g1688021; PIDN:BAAL1474.1; PID:g1688024

A:Experimental source: strain 168

C:Genetics:

A:Gene: ypeB

Query Match 22.9%; Score 55.5; DB 2; Length 450;  
Best Local Similarity 37.8%; Pred. No. 13;  
Matches 14; Conservative 4; Mismatches 18; Indels 1; Gaps 1;

QY 3 EALGSDYDPVCAVTLSTPLSAVREERKASDGSVPGF 39  
 : | | | | | | | | | | | | | | | | | | | |  
 DB 324 DKIGVFSYVPVENKVRYP-EAIRMKVALDDGGEVVG 359

## RESULT 15

G01946  
 nitric-oxide synthase (EC 1.14.13.39), neuronal - human  
 N:Alternate names: nitric oxide synthase 1  
 C:Species: Homo sapiens (man)  
 C:Date: 21-Dec-1996 #sequence,revision 06-Jun-1997 #text\_change 04-Mar-2000  
 C:Accession: G01946; I56508; S28878  
 R:Park, C.; Gianotti, C.; Park, R.; Krishna, G.  
 submitted to the EMBL Data Library, July 1995  
 A:Reference number: G08911  
 A:Accession: G01946  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-1433 <PAR>  
 A:Cross-references: EMBL:U31466; NID:9951318; PIDN:AAB49040.1; PID:9951319  
 R:Fujisawa, H.; Ogura, T.; Kurashima, Y.; Yokoyama, T.; Yamashita, J.; Esumi, H.  
 J. Neurochem. 63, 140-145, 1994  
 A:Title: Expression of two types of nitric oxide synthase mRNA in human neuroblastoma cell lines  
 A:Reference number: I56508; MUID:94267447  
 A:Accession: I56508  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-130, K', 132-177, 'LA', 179, 'RP', 182, 'G', 184-1433 <RES>  
 A:Cross-references: GB:D16408; NID:9506339; PIDN:BAA03895.1; PID:9987662  
 R:Nakane, M.; Schmidt, H.H.W.; Pollock, J.S.; Foerstermann, U.; Murad, F.  
 FEBS Lett. 316, 175-180, 1993  
 A:Title: Cloned human brain nitric oxide synthase is highly expressed in skeletal muscle  
 A:Reference number: S28878; MUID:93131039  
 A:Accession: S28878  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-130, K', 132-490, 'HR', 493-547, 'L', 549-561, 'A', 563-1362, 'I', 1364-1405, 'I', 1406-1433 <RES>  
 A:Cross-references: GB:L02881; NID:g189261  
 C:Genetics:  
 A:Gene: GDB:NOS1; NOS  
 A:Cross-references: GDB:132579; OMIM:163731  
 A:Map position: 12q24.2-12q24.31  
 C:Superfamily: GLGF domain homology; flavodoxin homology; NADPH--ferrihemoprotein reductase  
 C:Keywords: chromoprotein; flavoprotein; heme; iron; metalloprotein; oxidoreductase  
 F:23-95/Domain: GLGF domain homology <GLG>  
 F:759-1398/Domain: NADPH--ferrihemoprotein reductase homology <FEH>  
 F:761-939/Domain: flavodoxin homology <FLX>  
 F:419/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match 22.98; Score 55.5; DB 2; Length 1433;  
 Best Local Similarity 47.1%; Pred. No. 48;  
 Matches 16; Conservative 3; Mismatches 14; Indels 1; Gaps 1;

QY 11 PPVCAVTLSTPLSAVREERKASDG-SYPGFGQLH 43  
 : | | | | | | | | | | | | | | | | | | | |  
 DB 128 PPTAVDLSHQPPAGKEQPLAVDGSAGPGNGPQH 161

Search completed: June 1, 2001, 13:46:15  
 Job time: 57 sec





GenCore version 4.5

Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 1, 2001, 13:45:18 ; Search time 11.63 Seconds  
(without alignments)  
77.636 Million cell updates/sec

Title: US-09-331-723A-1

Perfect score: 242

Sequence: 1 AAALGSEFDYPPVGAATLSY.....ERKASDGVFGQLHPTQ 47

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 185757 seqs, 19210857 residues

Total number of hits satisfying chosen parameters: 185757

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: /cgn2\_6/ptodata/2/1aa/5A\_COMB.pep:\*\*

2: /cgn2\_6/ptodata/2/1aa/5B\_COMB.pep:\*\*

3: /cgn2\_6/ptodata/2/1aa/6A\_COMB.pep:\*\*

4: /cgn2\_6/ptodata/2/1aa/6B\_COMB.pep:\*\*

5: /cgn2\_6/ptodata/2/1aa/PCUTUS\_COMB.pep:\*\*

6: /cgn2\_6/ptodata/2/1aa/backfiles1.pep:\*\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	142.5	58.9	528	2	US-08-808-931-10
2	142.5	58.9	528	3	US-08-808-323-10
3	142.5	58.9	528	3	US-09-050-603A-10
4	142.5	58.9	528	3	US-09-102-420B-10
5	141.5	58.5	481	1	US-08-472-028A-6
6	141.5	58.5	481	2	US-08-808-931-6
7	141.5	58.5	481	3	US-08-808-323-6
8	141.5	58.5	481	3	US-09-050-603A-6
9	141.5	58.5	481	3	US-09-102-420B-6
10	141.5	58.5	483	4	US-09-071-296-6
11	140.5	58.1	312	2	US-08-808-931-22
12	140.5	58.1	312	3	US-08-808-323-22
13	140.5	58.1	312	3	US-09-050-603A-22
14	140.5	58.1	312	3	US-09-102-420B-22
15	139.5	57.6	539	2	US-08-808-931-16
16	139.5	57.6	539	3	US-08-808-323-16
17	139.5	57.6	539	3	US-09-050-603A-16
18	139.5	57.6	539	3	US-09-102-420B-16
19	138.5	57.2	543	2	US-08-808-931-12
20	138.5	57.2	543	3	US-08-808-323-12
21	138.5	57.2	543	3	US-09-050-603A-12
22	138.5	57.2	543	3	US-09-102-420B-12
23	137.5	56.8	440	2	US-08-808-931-24
24	137.5	56.8	440	3	US-08-808-323-24
25	137.5	56.8	440	3	US-09-050-603A-24
26	137.5	56.8	440	3	US-09-102-420B-24
27	133.5	55.2	537	1	US-08-472-028A-2

28	133.5	55.2	537	2	US-08-808-931-2	Sequence 2, Appli
29	133.5	55.2	537	3	US-08-808-323-2	Sequence 2, Appli
30	133.5	55.2	537	3	US-09-050-603A-2	Sequence 2, Appli
31	133.5	55.2	537	3	US-09-102-420B-2	Sequence 2, Appli
32	133.5	55.2	537	4	US-09-071-296-2	Sequence 2, Appli
33	132.5	54.8	560	2	US-08-808-931-18	Sequence 18, Appl
34	132.5	54.8	560	3	US-08-808-323-18	Sequence 18, Appl
35	132.5	54.8	560	3	US-09-050-603A-18	Sequence 18, Appl
36	132.5	54.8	560	3	US-09-102-420B-18	Sequence 18, Appl
37	131.5	54.3	536	2	US-08-808-931-20	Sequence 20, Appl
38	131.5	54.3	536	3	US-08-808-323-20	Sequence 20, Appl
39	131.5	54.3	536	3	US-09-050-603A-20	Sequence 20, Appl
40	131.5	54.3	536	3	US-09-102-420B-20	Sequence 20, Appl
41	57.5	23.8	1317	3	US-09-083-521-7	Sequence 7, Appli
42	57	23.6	982	2	US-08-673-789-4	Sequence 4, Appli
43	57	23.6	983	1	US-08-162-809-16	Sequence 16, Appl
44	57	23.6	983	1	US-08-167-919A-10	Sequence 10, Appl
45	57	23.6	983	2	US-08-449-645A-21	Sequence 21, Appl

## ALIGNMENTS

## RESULT 1

US-08-808-931-10

; Sequence 10, Application US/08808931

; Patent No. 5939602

; GENERAL INFORMATION:

; APPLICANT: Volrath, Sandra

; APPLICANT: Johnson, Marie

; APPLICANT: Potter, Sharon

; APPLICANT: Ward, Eric

; APPLICANT: Helfetz, Peter

; TITLE OF INVENTION: DNA Molecules Encoding Plant

; TITLE OF INVENTION: Protoporphyrinogen Oxidase and Inhibitor-Resistant Mutants

; TITLE OF INVENTION: Theoreof

; NUMBER OF SEQUENCES: 35

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: No. 5939602artis Corporation

; STREET: 520 White Plains Road, P.O. Box 2005

; CITY: Tarrytown

; STATE: NY

; COUNTRY: USA

; ZIP: 10591-9005

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/808,931

; FILING DATE:

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 60/012,705

; FILING DATE: 28-FEB-1996

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 60/013,612

; FILING DATE: 28-FEB-1996

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 60/020,003

; FILING DATE: 21-JUN-1996

; ATTORNEY/AGENT INFORMATION:

; NAME: Meligs, J. Timothy

; REGISTRATION NUMBER: 38,241

; REFERENCE/DOCKET NUMBER: CGC 1847

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (919) 541-8587

; TELEFAX: (919) 541-8689

; INFORMATION FOR SEQ ID NO: 10:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 528 amino acids

; TYPE: amino acid





COMPUTER: IBM PC compatible  
\* OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
FILING DATE: 21-JUN-1996  
ATTORNEY/AGENT INFORMATION:  
FILING DATE: 28-FEB-1996  
\* CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/012,705  
FILING DATE: 28-FEB-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/013,612  
FILING DATE: 28-FEB-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/020,003  
FILING DATE: 21-JUN-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Meigs, J. Timothy  
REGISTRATION NUMBER: 38,241  
REFERENCE/DOCKET NUMBER: CGC 1847  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (919) 541-8587  
TELEFAX: (919) 541-8689  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 481 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-808-931-6

Query Match 58.5%; Score 141.5; DB 2; Length 481;  
Best Local Similarity 61.7%; Pred. No. 3.8e-12;  
Matches 29; Conservative 6; Mismatches 11; Indels 1; Gaps 1;  
QY 1 AAALGSDYPPVAVTSLYPLSAVREERKASDGYPGQLHPRQ 47  
DB 297 AADALSRYPPVAAVTVSPKAIKKE-CLIDGELQGFGLHPRSQ 342

RESULT 7  
US-08-808-323-6  
Sequence 6, Application US/08808323  
Patent No. 6018105  
GENERAL INFORMATION:  
APPLICANT: Johnson, Marie  
APPLICANT: Volrath, Sandra  
APPLICANT: Ward, Eric  
TITLE OF INVENTION: Promoters from Plant  
NUMBER OF SEQUENCES: 26  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: No. 6018105artis Corporation  
STREET: 520 White Plains Road, P.O. Box 2005  
CITY: Tarrytown  
STATE: NY  
COUNTRY: USA  
ZIP: 10591-9005  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
FILING DATE: 28-FEB-1996  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/012,705  
FILING DATE: 28-FEB-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/013,612

FILING DATE: 28-FEB-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/020,003  
FILING DATE: 21-JUN-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Meigs, J. Timothy  
REGISTRATION NUMBER: 38,241  
REFERENCE/DOCKET NUMBER: CGC 1846  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (919) 541-8587  
TELEFAX: (919) 541-8689  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 481 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-808-323-6

Query Match 58.5%; Score 141.5; DB 3; Length 481;  
Best Local Similarity 61.7%; Pred. No. 3.8e-12;  
Matches 29; Conservative 6; Mismatches 11; Indels 1; Gaps 1;  
QY 1 AAALGSDYPPVAVTSLYPLSAVREERKASDGYPGQLHPRQ 47  
DB 297 AADALSRYPPVAAVTVSPKAIKKE-CLIDGELQGFGLHPRSQ 342

RESULT 8  
US-09-050-603A-6  
Sequence 6, Application US/09050603A  
Patent No. 6023012  
GENERAL INFORMATION:  
APPLICANT: Volrath, Sandra  
APPLICANT: Johnson, Marie  
APPLICANT: Potter, Sharon  
APPLICANT: Ward, Eric  
APPLICANT: Heifetz, Peter  
TITLE OF INVENTION: DNA Molecules Encoding Plant  
NUMBER OF SEQUENCES: 37  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: No. 6023012artis Corporation  
STREET: 3054 Cornwallis Road  
CITY: Research Triangle Park  
STATE: NC  
COUNTRY: USA  
ZIP: 27709  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
FILING DATE: 30-MAR-1998  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/808,931  
FILING DATE: 28-FEB-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/012,705  
FILING DATE: 28-FEB-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/013,612  
FILING DATE: 28-FEB-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/020,003  
FILING DATE: 21-JUN-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Meigs, J. Timothy  
REGISTRATION NUMBER: 38,241

REFERENCE/DOCKET NUMBER: CGC 1847  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (919) 541-8587  
TELEFAX: (919) 541-8689  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 481 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-050-603A-6

Query Match 58.5%; Score 141.5; DB 3; Length 481;  
Best Local Similarity 61.7%; Pred. No. 3.8e-12;  
Matches 29; Conservative 6; Mismatches 11; Indels 1; Gaps 1;

QY 1 AAEALGSDYPPVGVAVTSLSAVREERKASDGSVPFGQLHPRQ 47  
DB 297 AADALSRFYPPVAAVTSYPKEAIRKE-CLIDGELQFGQLHPRSQ 342

## RESULT 9

US-09-102-420B-6  
Sequence 6, Application US/09102420B  
Patent No. 6084155  
GENERAL INFORMATION:  
APPLICANT: Volrath, Sandra  
APPLICANT: Johnson, Marie  
APPLICANT: Ward, Eric  
APPLICANT: Heifetz, Peter  
TITLE OF INVENTION: HERBICIDE-TOLERANT PROTOPORPHYRINOGEN  
NUMBER OF SEQUENCES: 43  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: No. 6084155artis Corporation  
STREET: 3054 Cornwallis Road  
CITY: Research Triangle Park  
STATE: NC USA  
COUNTRY: USA  
ZIP: 27709  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/102,420B  
FILING DATE: 22-JUN-1998  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 09/059,164  
FILING DATE: 13-APR-1998  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 09/050,603  
FILING DATE: 30-MAR-1998  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/126,430  
FILING DATE: 11-MAR-1998  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/808,931  
FILING DATE: 28-FEB-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/012,705  
FILING DATE: 28-FEB-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/013,612  
FILING DATE: 28-FEB-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/020,003  
FILING DATE: 21-JUN-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/472,028

FILING DATE: 06-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Meigs, J. Timothy  
REGISTRATION NUMBER: 38,241  
REFERENCE/DOCKET NUMBER: CGC 1847/CIP3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (919) 541-8587  
TELEFAX: (919) 541-8689  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 481 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-102-420B-6

Query Match 58.5%; Score 141.5; DB 3; Length 481;  
Best Local Similarity 61.7%; Pred. No. 3.8e-12;  
Matches 29; Conservative 6; Mismatches 11; Indels 1; Gaps 1;

QY 1 AAEALGSDYPPVGVAVTSLSAVREERKASDGSVPFGQLHPRQ 47  
DB 297 AADALSRFYPPVAAVTSYPKEAIRKE-CLIDGELQFGQLHPRSQ 342

## RESULT 10

US-09-071-296-6  
Sequence 6, Application US/09071296  
Patent No. 6177245  
GENERAL INFORMATION:  
APPLICANT: Ward, Eric R  
APPLICANT: Volrath, Sandra  
TITLE OF INVENTION: Manipulation of Protoporphyrinogen  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Ciba-Geigy Corporation  
STREET: 7 Skyline Drive  
CITY: Hawthorne  
STATE: NY  
COUNTRY: USA  
ZIP: 10532  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/071,296  
FILING DATE: 06-JUN-1995  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/261,198  
FILING DATE: 16-JUN-94  
ATTORNEY/AGENT INFORMATION:  
NAME: Elmer, James Scott  
REGISTRATION NUMBER: 36,129  
REFERENCE/DOCKET NUMBER: CGC 1748/CIP  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 919-541-8614  
TELEFAX: 919-541-8689  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 483 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-071-296-6

Query Match 58.5%; Score 141.5; DB 4; Length 483;  
Best Local Similarity 61.7%; Pred. No. 3.8e-12;



NUMBER OF SEQUENCES: 37  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: No. 603012artlis Corporation  
STREET: 3054 Cornwallis Road  
CITY: Research Triangle Park  
STATE: NC  
COUNTRY: USA  
ZIP: 27709  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/050,603A  
FILING DATE: 30-MAR-1998  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/808,931  
FILING DATE: 28-FEB-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/012,705  
FILING DATE: 28-FEB-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/013,612  
FILING DATE: 28-FEB-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/020,003  
FILING DATE: 21-JUN-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Meigs, J. Timothy  
REGISTRATION NUMBER: 38,241  
REFERENCE/DOCKET NUMBER: CGC 1847  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (919) 541-8587  
TELEFAX: (919) 541-8689  
INFORMATION FOR SEQ ID NO: 22:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 312 amino acids  
TYPE: amino acid  
STRANDEDNESS: not relevant  
TOPOLOGY: not relevant  
MOLECULE TYPE: protein  
US-09-050-603A-22

Query Match 58.1%; Score 140.5; DB 3; Length 312;  
Best Local Similarity 61.7%; Pred. No. 3e-12;  
Matches 29; Conservative 6; Mismatches 11; Indels 1; Gaps 1;

QY 1 AAALGSGFDYPPVAVNTLSYPLSAVREERKASDGSVPFGQLHPRQ 47  
DB 128 ADALSLFYPPVAANTVSPKEAIRKE-CLIDGELQGFGLHPRSQ 173

RESULT 14  
US-09-102-420B-22  
Sequence 22, Application US/09102420B  
Patent No. 6084155  
GENERAL INFORMATION:  
APPLICANT: Volrath, Sandra  
APPLICANT: Johnson, Marie  
APPLICANT: Ward, Eric  
APPLICANT: Helfetz, Peter  
TITLE OF INVENTION: HERBICIDE-TOLERANT PROTOPORPHYRINOGEN  
TITLE OF INVENTION: OXIDASE ("PROTOX")  
NUMBER OF SEQUENCES: 43  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: No. 6084155artlis Corporation  
STREET: 3054 Cornwallis Road  
CITY: Research Triangle Park  
STATE: NC  
COUNTRY: USA

ZIP: 27709  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/102,420B  
FILING DATE: 22-JUN-1998  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 09/059,164  
FILING DATE: 13-APR-1998  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 09/050,603  
FILING DATE: 30-MAR-1998  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/126,430  
FILING DATE: 11-MAR-1998  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/808,931  
FILING DATE: 28-FEB-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/012,705  
FILING DATE: 28-FEB-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/013,612  
FILING DATE: 28-FEB-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/020,003  
FILING DATE: 21-JUN-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/472,028  
FILING DATE: 06-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Meigs, J. Timothy  
REGISTRATION NUMBER: 38,241  
REFERENCE/DOCKET NUMBER: CGC 1847/CIP3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (919) 541-8587  
TELEFAX: (919) 541-8689  
INFORMATION FOR SEQ ID NO: 22:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 312 amino acids  
TYPE: amino acid  
STRANDEDNESS: not relevant  
TOPOLOGY: not relevant  
MOLECULE TYPE: protein  
US-09-102-420B-22

Query Match 58.1%; Score 140.5; DB 3; Length 312;  
Best Local Similarity 61.7%; Pred. No. 3e-12;  
Matches 29; Conservative 6; Mismatches 11; Indels 1; Gaps 1;

QY 1 AAALGSGFDYPPVAVNTLSYPLSAVREERKASDGSVPFGQLHPRQ 47  
DB 128 ADALSLFYPPVAANTVSPKEAIRKE-CLIDGELQGFGLHPRSQ 173

RESULT 15  
US-08-808-931-16  
Sequence 16, Application US/08808931  
Patent No. 5939602  
GENERAL INFORMATION:  
APPLICANT: Volrath, Sandra  
APPLICANT: Johnson, Marie  
APPLICANT: Potter, Sharon  
APPLICANT: Ward, Eric  
APPLICANT: Helfetz, Peter  
TITLE OF INVENTION: DNA Molecules Encoding Plant  
TITLE OF INVENTION: Protoporphyrinogen Oxidase and Inhibitor-Resistant Mutants  
TITLE OF INVENTION: Thereof

NUMBER OF SEQUENCES: 35  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: NO. 5939602artis Corporation  
STREET: 520 White Plains Road, P.O. Box 2005  
CITY: Tarrytown  
STATE: NY  
COUNTRY: USA  
ZIP: 10591-9005  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/808,931  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/012,705  
FILING DATE: 28-FEB-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/013,612  
FILING DATE: 28-FEB-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/020,003  
FILING DATE: 21-JUN-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Meigs, J. Timothy  
REGISTRATION NUMBER: 38,241  
REFERENCE/DOCKET NUMBER: CGC 1847  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (919) 541-8587  
TELEFAX: (919) 541-8689  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 539 amino acids  
TYPE: amino acid  
STRANDEDNESS: not relevant  
TOPOLOGY: not relevant  
MOLECULE TYPE: protein  
US-08-808-931-16

Query Match 57.6%; Score 139.5; DB 2; Length 539;  
Best Local Similarity 59.6%; Pred. No. 8.4e-12;  
Matches 28; Conservative 7; Mismatches 11; Indels 1; Gaps 1;  
Qy 1 AAEALGSFDYPPVGVAVTILSYPLSAVREERKASDGSVPFGQLHPTQ 47  
Db 355 AADALSQFYPPVAVSVTVSKAIRRE-CLIDGELKGFQQLHPRSQ 400  
||:| | ||| :||| | :| | || :||| | |  
| :| | | | :| | | | :| | | | |

Search completed: June 1, 2001, 13:45:57  
Job time: 39 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 1, 2001, 13:45:18 ; Search time 18.55 Seconds  
(without alignments)  
144.834 Million cell updates/sec

Title: US-09-331-723a-1

Perfect score: 242  
Sequence: 1 AAEALGSDYPPVGVATLSY.....ERRASDGVPGFGLHPRQTQ 47

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 390729 seqs, 57163235 residues

Total number of hits satisfying chosen parameters: 390729

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_0401:\*

1: /SID56/gcgdata/geneseq/geneseq/AA1980.DAT:\*  
2: /SID56/gcgdata/geneseq/geneseq/AA1981.DAT:\*  
3: /SID56/gcgdata/geneseq/geneseq/AA1982.DAT:\*  
4: /SID56/gcgdata/geneseq/geneseq/AA1983.DAT:\*  
5: /SID56/gcgdata/geneseq/geneseq/AA1984.DAT:\*  
6: /SID56/gcgdata/geneseq/geneseq/AA1985.DAT:\*  
7: /SID56/gcgdata/geneseq/geneseq/AA1986.DAT:\*  
8: /SID56/gcgdata/geneseq/geneseq/AA1987.DAT:\*  
9: /SID56/gcgdata/geneseq/geneseq/AA1988.DAT:\*  
10: /SID56/gcgdata/geneseq/geneseq/AA1989.DAT:\*  
11: /SID56/gcgdata/geneseq/geneseq/AA1990.DAT:\*  
12: /SID56/gcgdata/geneseq/geneseq/AA1991.DAT:\*  
13: /SID56/gcgdata/geneseq/geneseq/AA1992.DAT:\*  
14: /SID56/gcgdata/geneseq/geneseq/AA1993.DAT:\*  
15: /SID56/gcgdata/geneseq/geneseq/AA1994.DAT:\*  
16: /SID56/gcgdata/geneseq/geneseq/AA1995.DAT:\*  
17: /SID56/gcgdata/geneseq/geneseq/AA1996.DAT:\*  
18: /SID56/gcgdata/geneseq/geneseq/AA1997.DAT:\*  
19: /SID56/gcgdata/geneseq/geneseq/AA1998.DAT:\*  
20: /SID56/gcgdata/geneseq/geneseq/AA1999.DAT:\*  
21: /SID56/gcgdata/geneseq/geneseq/AA2000.DAT:\*  
22: /SID56/gcgdata/geneseq/geneseq/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	242	100.0	47	W64450	C. reinhardtii por
2	242	100.0	563	Y49538	Chlamydomonas rein
3	242	100.0	563	Y54101	An algal protoporp
4	164.5	68.0	548	Y05200	N. tabacum protopo
5	164.5	68.0	548	Y05201	N. tabacum protopo
6	142.5	58.9	528	W41607	Wheat prototox-1. F
7	142.5	58.9	528	W25738	Wheat protoporphyr
8	141.5	58.5	46	W64452	Z. mays porphyrin
9	141.5	58.5	481	W41605	Maize prototox-1. Z
10	141.5	58.5	481	W25745	Maize protoporphyr
11	141.5	58.5	481	W51255	Maize proto-porphyr

12	141.5	58.5	483	17	R90297
13	140.5	58.1	312	18	W41612
14	140.5	58.1	312	18	W25743
15	139.5	57.6	539	18	W41608
16	139.5	57.6	539	18	W41609
17	139.5	57.6	539	18	W25740
18	138.5	57.2	543	18	W25739
19	138.5	57.2	543	20	W92498
20	138.5	57.2	543	21	Y54100
21	137.5	56.8	440	18	W41613
22	137.5	56.8	440	18	W25744
23	133.5	55.2	46	19	W64451
24	133.5	55.2	413	21	G42387
25	133.5	55.2	416	21	G42386
26	133.5	55.2	537	17	R90295
27	133.5	55.2	537	18	W41603
28	133.5	55.2	537	18	W25746
29	133.5	55.2	537	19	W51347
30	133.5	55.2	537	21	G42385
31	132.5	54.8	560	18	W41610
32	132.5	54.8	560	18	W25741
33	131.5	54.3	536	18	W41611
34	131.5	54.3	536	18	W25742
35	57	23.6	983	14	R31466
36	57	23.6	983	16	R75711
37	57	23.6	983	21	B08665
38	57	23.6	983	21	B08666
39	56	23.1	346	18	W21730
40	56	23.1	363	18	W23658
41	56	23.1	363	19	W75699
42	56	23.1	363	19	W68185
43	56	23.1	379	21	B58841
44	55.5	22.9	1003	21	B24058
45	55.5	22.9	1004	21	Y66639

#### ALIGNMENTS

RESULT 1  
W64450  
ID W64450 standard; Protein; 47 AA.  
XX AC W64450;  
XX 16-OCT-1998 (first entry)  
XX C. reinhardtii porphyrin herbicide resistance domain protein.  
XX DE Herbicide resistance; porphyrin; protoporphyrinogen oxidase; PPO;  
XX KW crop plant; selective weed control.  
XX OS Chlamydomonas reinhardtii.  
XX PN W09829554-AA.  
XX PD 09-JUL-1998.  
XX PF 27-DEC-1996; 96WO-US20415.  
XX PR 27-DEC-1996; 96WO-US20415.  
XX (SUMO) SUMITOMO CHEM CO LTD.  
XX (UYDU-) UNIV DUKE.  
XX Boynton JE, Gillham NW, Ishige F, Randolph-Anderson BL;  
XX Sato R;  
XX WPI; 1998-388134/33.  
XX DR N-PSDB; V46291.  
XX Conferring resistance to herbicides that inhibit proto-porphyrinogen  
XX oxidase - by introducing DNA that encodes a resistant form of the

PT enzyme, also resistant plants and algae, mutant DNA, vectors and  
PT transformed microorganisms

PS Claim 1; Page 65; 108pp; English.

XX This sequence represents a fragment of the Chlamydomonas reinhardtii  
CC porphyrin herbicide resistance domain which is used in a method to  
CC investigate resistance to herbicides that inhibit protoporphyrinogen  
CC oxidase (PPO). The method is applied to crop plants, allowing use of  
CC herbicides for selective weed control. The method can be used to  
CC identify compounds that are, or are not, PPO-inhibitors.

XX Sequence 47 AA;

Query Match 100.0%; Score 242; DB 19; Length 47;  
Best Local Similarity 100.0%; Pred. No. 8.3e-27;  
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAALGSDYPPVPGAVTSLYPLSAVREERKASDGVFGQLHPTQ 47

Db 1 aaalgsfdyppvgavtlyplsavreerkasdgsvpgfqlhprtq 47

RESULT 2

Y49538  
ID Y49538 standard; Protein: 563 AA.

XX Y49538;

XX 12-JAN-2000 (first entry)

XX Chlamydomonas reinhardtii protoporphyrinogen oxidase.

DE Protoporphyrinogen oxidase; PPO; herbicide; inhibition; resistance;  
XX malignant cell growth; phototropic herbicide.

XX Chlamydomonas reinhardtii.

XX EP955380-A2.

XX 10-NOV-1999.

XX 09-APR-1999; 99EP-0107037.

XX 10-APR-1998; 98JP-0099619.

XX (SUMO) SUMITOMO CHEM CO LTD.

XX Shimokawatoko Y, Nishio S;

XX WPI; 1999-612726/53.

DR N-PSDB; 232119.

XX Evaluating ability of compounds to inhibit protoporphyrinogen oxidase  
PT activity for development of herbicides -

PS Claim 19; Page 26-28; 46pp; English.

XX A method has been developed for evaluating the ability of a compound to  
CC inhibit protoporphyrinogen oxidase (PPO) activity. The method comprises:  
CC (a) culturing a transformant expressing a protoporphyrinogen oxidase  
CC gene in the presence and absence of a test compound to measure a growth  
CC rate; and (b) determining the ability of the compound to inhibit PPO  
CC activity by comparing growth rates. The method is useful for the  
CC development of PPO inhibiting-type herbicides. Also the method is useful  
CC for screening medications having the ability for inhibiting the  
CC malignant cell growth in addition to evaluating the potency of a  
CC phototropic herbicide, screening a compound having a herbicidal  
CC activity. The PPO gene can be used to obtain or generate a gene  
CC resistance to a PPO activity inhibiting agent e.g. a phototropic  
CC herbicide. The present sequence represents Chlamydomonas reinhardtii-  
CC derived PPO.

XX Sequence 563 AA;

Query Match 100.0%; Score 242; DB 20; Length 563;  
Best Local Similarity 100.0%; Pred. No. 1.5e-25;  
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAALGSDYPPVPGAVTSLYPLSAVREERKASDGVFGQLHPTQ 47

Db 377 aaalgsfdyppvgavtlyplsavreerkasdgsvpgfqlhprtq 423

RESULT 3

Y54101

ID Y54101 standard; Protein: 563 AA.

XX Y54101;

XX 27-MAR-2000 (first entry)

XX An algal protoporphyrin IX oxidase protein.

DE Weed control; weed activity; plant; protoporphyrin IX oxidase; PPO;  
KW herbicide; chlormethoxynil; acifluorfen; protoporphyrin X.

XX Chlamydomonas reinhardtii.

XX EP953646-A2.

XX 03-NOV-1999.

XX 30-APR-1999; 99EP-0108463.

XX 30-APR-1998; 98JP-0120553.

XX 02-OCT-1998; 98JP-0281127.

XX 20-NOV-1998; 98JP-0330981.

XX 02-MAR-1999; 99JP-0054730.

XX (SUMO) SUMITOMO CHEM CO LTD.

XX Nakajima H, Nagasawa A;

XX WPI; 2000-088762/08.

DR N-PSDB; 245282.

XX Method of making plants resistant to weed control compounds by  
PT introducing gene encoding protein into plant cell -

PS Example 15; Page 77-79; 119pp; English.

XX The present sequence represents an algal protoporphyrin IX oxidase (PPO).  
CC The sequence was used in the course of the invention. The specification  
CC describes a method of making plants resistant to weed control compounds.  
CC The method comprises introducing a gene encoding a protein into a plant  
CC cell, where the protein has a specific affinity for a substance which  
CC is concerned with the weed activity of a weed control compound,  
CC e.g. protoporphyrin IX; has no ability to modify this protein; and  
CC is virtually free from framework regions in an immunoglobulin; and  
CC expressing the gene. The gene of the invention is used for producing a  
CC plant resistant to specified compounds, such as protoporphyrin IX  
CC oxidase (PPO) inhibitory-type herbicides, e.g. chlormethoxynil,  
CC acifluorfen, etc.. When PPO is inhibited, protoporphyrin IX (the  
CC substrate of PPO) accumulates in the plant cells, and is metabolised to  
CC form protoporphyrin X, followed by formation of active oxygen in the  
CC presence of protoporphyrin X and light, which damages cell functions.

XX Sequence 563 AA;

Query Match 100.0%; Score 242; DB 21; Length 563;  
Best Local Similarity 100.0%; Pred. No. 1.5e-25;  
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



XX The present sequence is wheat protoporphyrinogen oxidase-1  
 CC (protox-1).  
 CC The protox-1 promoter can be used to express herbicide resistant  
 CC enzymes, specifically protox, i.e. a plant tissue, plant or progeny  
 CC containing a chimeric gene of the promoter and a heterologous  
 CC coding sequence. The plant can also be used in breeding programmes.  
 CC Also hybridising fragments of the protox coding sequence can be  
 CC used as probes, e.g. to isolate related genes or for genomic  
 CC mapping.  
 XX sequence 528 AA;  
 SQ  
 Query Match 58.9%; Score 142.5; DB 18; Length 528;  
 Best Local Similarity 61.7%; Pred. No. 1.1e-11;  
 Matches 29; Conservative 6; Mismatches 11; Indels 1; Gaps 1;  
 QY 1 AAALGSDYPPVGAVTSLYPLSAVREERKASDGSVPFGQLHPRQ 47  
 Db 344 aadalskfyppvaavtvsypkeairke-clidgelgfgqlhprsq 389  
 RESULT 7  
 W25738  
 ID W25738 standard; Protein: 528 AA.  
 AC W25738;  
 XX  
 XX 01-MAR-1998 (first entry)  
 DE Wheat protoporphyrinogen oxidase (protox-1).  
 XX  
 XX Protox-1a; protoporphyrinogen oxidase; inhibitor; wheat;  
 KW herbicide tolerance; herbicide resistance; transgenic plant.  
 XX  
 OS Triticum aestivum cv. Kanzler.  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 211 /note= "substitution of Ala-211 by another amino  
 FT acid, especially Val or Thr, provides a  
 FT modified protox tolerant to a herbicide  
 FT (Claims 37-38)"  
 FT  
 FT Misc-difference 212 /note= "substitution of Gly-212 by another amino  
 FT acid, especially Ser, provides a modified  
 FT protox tolerant to a herbicide (Claims  
 FT 39-40)"  
 FT  
 FT Misc-difference 356 /note= "substitution of Val-356 by another amino  
 FT acid, especially Leu, provides a modified  
 FT protox tolerant to a herbicide (Claims  
 FT 31-32)"  
 FT  
 FT Misc-difference 421 /note= "substitution of Ser-421 by another amino  
 FT acid, especially Pro, provides a modified  
 FT protox tolerant to a herbicide (Claims  
 FT 33-34)"  
 FT  
 FT Misc-difference 466 /note= "substitution of Ile-466 by another amino  
 FT acid, especially Thr, provides a modified  
 FT protox tolerant to a herbicide (Claims  
 FT 41-42)"  
 FT  
 FT Misc-difference 502 /note= "substitution of Val-502 by another amino  
 FT acid, especially Ala, provides a modified  
 FT protox tolerant to a herbicide (Claims  
 FT 35-36)"  
 FT  
 XX W09732011-A1.  
 PN  
 XX 04-SEP-1997.  
 PD

XX 27-FEB-1997; 97WO-US03313.  
 XX  
 XX 21-JUN-1996; 96US-0020003.  
 PR 28-FEB-1996; 96US-0012705.  
 PR 28-FEB-1996; 96US-0013612.  
 XX (NOVS ) NOVARTIS AG.  
 PA  
 XX Heifetz PB, Johnson MA, Potter SL, Volrath SL, Ward ER;  
 PI WPI; 1997-448683/41.  
 XX N-PSDB; T86121.  
 DR  
 XX New DNA encoding plant protoporphyrinogen oxidase enzyme - and  
 PT herbicide resistant mutants, useful to prepare plants resistant to  
 PT herbicide which therefore kills undesired vegetation only  
 XX  
 PS Claim 3; Page 120-123; 196pp; English.  
 XX  
 XX This protein comprises wheat protoporphyrinogen oxidase (protox-1),  
 CC an enzyme that catalyses the oxidation of protoporphyrinogen IX to  
 CC protoporphyrin IX. Its amino acid sequence was deduced from a  
 CC Protox-1 cDNA clone (see T86121). Sites within the wheat Protox-1  
 CC gene have been identified that can be mutated to encode a modified  
 CC protox that is resistant to protox inhibitors and hence tolerant of  
 CC certain herbicides. Plants, especially crop plants, may be  
 CC engineered for resistance to protox inhibitors via mutation of the  
 CC native protox gene to a resistant form, or they may be transformed  
 CC with a gene encoding an inhibitor-resistant form of a plant protox  
 CC enzyme, such as claimed forms from wheat, soybean, cotton,  
 CC sugarbeet, oilseed rape, rice and sorghum (see W25738-48).  
 CC Application of herbicide will then kill undesired vegetation only.  
 CC Protox enzymes can also be expressed in transformed host cells and  
 CC used to identify inhibitors of protox enzyme activity, i.e.  
 CC herbicide candidates, or to design herbicide tolerant forms of the  
 CC enzyme.  
 XX  
 XX Sequence 528 AA;  
 SQ  
 Query Match 58.9%; Score 142.5; DB 18; Length 528;  
 Best Local Similarity 61.7%; Pred. No. 1.1e-11;  
 Matches 29; Conservative 6; Mismatches 11; Indels 1; Gaps 1;  
 QY 1 AAALGSDYPPVGAVTSLYPLSAVREERKASDGSVPFGQLHPRQ 47  
 Db 344 aadalskfyppvaavtvsypkeairke-clidgelgfgqlhprsq 389  
 RESULT 8  
 W64452  
 ID W64452 standard; Protein: 46 AA.  
 XX  
 XX W64452;  
 XX  
 XX 16-OCT-1998 (first entry)  
 DT  
 XX Z. mays porphyrin herbicide resistance domain protein.  
 DE  
 XX  
 KW Herbicide resistance; porphyrin; protoporphyrinogen oxidase; PPO;  
 KW crop plant; selective weed control; corn.  
 XX  
 OS Zea mays.  
 XX  
 XX W09829554-A1.  
 PN  
 XX 09-JUL-1998.  
 PD  
 XX 27-DEC-1996; 96WO-US20415.  
 PF  
 XX 27-DEC-1996; 96WO-US20415.  
 XX  
 PD

PA (SUMO ) SUMITOMO CHEM CO LTD.  
 XX (YUDU-) UNIV DUKE.  
 PI Boynton JE, Gillham NW, Ishige F, Randolph-Anderson BL;  
 XX Sato R;  
 DR WPI; 1998-388134/33.  
 DR N-PSDB; V46293.  
 XX Conferring resistance to herbicides that inhibit proto-porphyrinogen  
 PT oxidase - by introducing DNA that encodes a resistant form of the  
 PT enzyme, also resistant plants and algae, mutant DNA, vectors and  
 PT transformed microorganisms  
 XX  
 PS Claim 1: Page 66; 108pp; English.  
 CC This sequence represents a fragment of the Zea mays strain B73 Inbred  
 CC porphyrin herbicide resistance domain which is used in a method to  
 CC investigate resistance to herbicides that inhibit protoporphyrinogen  
 CC oxidase (PPO). The method is applied to crop plants, allowing use of  
 CC herbicides for selective weed control. The method can be used to  
 CC identify compounds that are, or are not, PPO-inhibitors.  
 XX  
 SQ Sequence 46 AA;  
 Query Match 58.5%; Score 141.5; DB 19; Length 46;  
 Best Local Similarity 61.7%; Pred. No. 9.2e-13;  
 Matches 29; Conservative 6; Mismatches 11; Indels 1; Gaps 1;  
 QY 1 AAPALGSFDYPPVGAVTSLYSPLSAVREERKASDGSVPFGQLHPRTQ 47  
 DB 1 aadalsrfyppvaavtvsypkeairke-clidgelqgfgqlhprsq 46  
 RESULT 9  
 ID W41605 standard; Protein; 481 AA.  
 XX  
 AC W41605;  
 XX  
 DT 20-APR-1998 (first entry)  
 XX  
 DE Maize protox-1.  
 XX  
 KW Protoporphyrinogen oxidase-1; protox-1; promoter; maize;  
 KW herbicide resistance; breeding programme; probe; gene isolation;  
 KW genomic mapping.  
 XX  
 OS Zea mays.  
 XX  
 PN W09732028-A1.  
 XX  
 PD 04-SEP-1997.  
 XX  
 PF 27-FEB-1997; 97WO-US03343.  
 XX  
 PR 21-JUN-1996; 96US-0020003.  
 PR 28-FEB-1996; 96US-0012705.  
 PR 28-FEB-1996; 96US-0013612.  
 XX  
 PA (NOVS ) NOVARTIS AG.  
 XX  
 PI Johnson MA, Vollrath SL, Ward ER;  
 XX  
 DR WPI; 1997-489209/45.  
 DR N-PSDB; V04306.  
 XX  
 PT DNA containing a plant proto-porphyrinogen oxidase gene promoter -  
 PT optionally linked to a heterologous gene, especially to express  
 PT herbicide-resistant enzymes, and plants containing such constructs  
 XX  
 PS Claim 29; Pages 49-52; 114pp; English.

XX The present sequence is maize protoporphyrinogen oxidase-1  
 CC (protox-1).  
 CC The protox-1 promoter can be used to express herbicide resistant  
 CC enzymes, specifically protox, i.e. a plant tissue, plant or progeny  
 CC containing a chimeric gene of the promoter and a heterologous  
 CC coding sequence. The plant can also be used in breeding programmes.  
 CC Also hybridising fragments of the protox coding sequence can be  
 CC used as probes, e.g. to isolate related genes or for genomic  
 CC mapping.  
 XX  
 SQ Sequence 481 AA;  
 Query Match 58.5%; Score 141.5; DB 18; Length 481;  
 Best Local Similarity 61.7%; Pred. No. 1.4e-11;  
 Matches 29; Conservative 6; Mismatches 11; Indels 1; Gaps 1;  
 QY 1 AAPALGSFDYPPVGAVTSLYSPLSAVREERKASDGSVPFGQLHPRTQ 47  
 DB 297 aadalsrfyppvaavtvsypkeairke-clidgelqgfgqlhprsq 342  
 RESULT 10  
 ID W25745 standard; Protein; 481 AA.  
 XX  
 AC W25745;  
 XX  
 DT 01-MAR-1998 (first entry)  
 XX  
 DE Maize protoporphyrinogen oxidase (protox-1).  
 XX  
 KW Protox-1; protoporphyrinogen oxidase; inhibitor; maize;  
 KW herbicide tolerance; herbicide resistance; transgenic plant.  
 XX  
 OS Zea mays.  
 XX  
 FH Key  
 FT Misc-difference 159  
 FT /note= "substitution of Cys-159 by another amino  
 FT acid, especially Phe or Lys, provides a  
 FT modified protox tolerant to a herbicide  
 FT (Claims 19-20)"  
 FT  
 FT Misc-difference 164  
 FT /note= "substitution of Ala-164 by another amino  
 FT acid, especially Thr, Leu or Val, provides  
 FT a modified protox tolerant to a herbicide  
 FT (Claims 25-26)"  
 FT  
 FT Misc-difference 165  
 FT /note= "substitution of Gly-165 by another amino  
 FT acid, especially Ser or Leu, provides a  
 FT modified protox tolerant to a herbicide  
 FT (Claims 27-28)"  
 FT  
 FT Misc-difference 370  
 FT /note= "substitution of Tyr-370 by another amino  
 FT acid, especially Ile or Met, provides a  
 FT modified protox tolerant to a herbicide  
 FT (Claims 29-30)"  
 FT  
 FT Misc-difference 419  
 FT /note= "substitution of Ile-419 by another amino  
 FT acid, especially Thr, His, Gly or Asn  
 FT provides a modified protox tolerant to a  
 FT herbicide (Claims 22-23)"  
 FT  
 XX W09732011-A1.  
 XX  
 PD 04-SEP-1997.  
 XX  
 PF 27-FEB-1997; 97WO-US03313.  
 XX  
 PR 21-JUN-1996; 96US-0020003.  
 PR 28-FEB-1996; 96US-0012705.





1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
84







GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 1, 2001, 23:57:42 ; Search time 1131.07 Seconds  
(without alignments)  
1089.048 Million cell updates/sec

Title: US-09-331-723A-4

Perfect score: 141

Sequence: 1 gccgcgagccctgggtc.....agctgacccgcgcgcag 141

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 9623517 seqs, 4368049070 residues

Total number of hits satisfying chosen parameters: 19247034

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*  
1: gb\_est1.\*  
2: gb\_est2.\*  
3: gb\_est3.\*  
4: gb\_est4.\*  
5: gb\_est5.\*  
6: gb\_est6.\*  
7: gb\_est7.\*  
8: gb\_est8.\*  
9: gb\_est9.\*  
10: gb\_est10.\*  
11: gb\_est11.\*  
12: gb\_est12.\*  
13: gb\_est13.\*  
14: gb\_est14.\*  
15: gb\_est15.\*  
16: gb\_est16.\*  
17: gb\_est17.\*  
18: gb\_est18.\*  
19: gb\_est19.\*  
20: gb\_est20.\*  
21: gb\_est21.\*  
22: gb\_est22.\*  
23: gb\_est23.\*  
24: gb\_est24.\*  
25: gb\_est25.\*  
26: gb\_est26.\*  
27: gb\_est27.\*  
28: gb\_est28.\*  
29: gb\_est29.\*  
30: gb\_est30.\*  
31: gb\_est31.\*  
32: gb\_est32.\*  
33: gb\_est33.\*  
34: gb\_est34.\*  
35: gb\_est35.\*  
36: gb\_est36.\*  
37: gb\_est37.\*  
38: gb\_est38.\*  
39: gb\_est39.\*  
40: gb\_est40.\*  
41: gb\_est41.\*  
42: gb\_est42.\*  
43: gb\_est43.\*  
44: gb\_est44.\*  
45: gb\_est45.\*  
46: gb\_est46.\*  
47: gb\_est47.\*

44: em\_esthum10.\*  
45: em\_esthum11.\*  
46: em\_esthum12.\*  
47: em\_esthum13.\*  
48: em\_esthum14.\*  
49: em\_esthum15.\*  
50: em\_esthum16.\*  
51: em\_esthum17.\*  
52: em\_esthum18.\*  
53: em\_esthum19.\*  
54: em\_esthum20.\*  
55: em\_esthum21.\*  
56: em\_esthum22.\*  
57: em\_esthum23.\*  
58: em\_esthum24.\*  
59: em\_esthum25.\*  
60: em\_esthum26.\*  
61: em\_esthum27.\*  
62: em\_esthum28.\*  
63: em\_estin1.\*  
64: em\_estin2.\*  
65: em\_estin3.\*  
66: em\_estin4.\*  
67: em\_estin5.\*  
68: em\_estom1.\*  
69: em\_estom2.\*  
70: em\_estov1.\*  
71: em\_estov2.\*  
72: em\_estp11.\*  
73: em\_estp12.\*  
74: em\_estp13.\*  
75: em\_estp14.\*  
76: em\_estp15.\*  
77: em\_estp16.\*  
78: em\_estp17.\*  
79: em\_estp18.\*  
80: em\_estp19.\*  
81: em\_estp10.\*  
82: em\_estro1.\*  
83: em\_estro2.\*  
84: em\_estro3.\*  
85: em\_estro4.\*  
86: em\_estro5.\*  
87: em\_estro6.\*  
88: em\_estro7.\*  
89: em\_estro8.\*  
90: em\_estro9.\*  
91: em\_estro10.\*  
92: em\_estro11.\*  
93: em\_estro12.\*  
94: em\_estro13.\*  
95: em\_estro14.\*  
96: em\_estro15.\*  
97: em\_estro16.\*  
98: em\_estro17.\*  
99: em\_estro18.\*  
100: em\_estro19.\*  
101: em\_estro20.\*  
102: gb\_est25.\*  
103: gb\_est26.\*  
104: gb\_est27.\*  
105: gb\_est28.\*  
106: gb\_est29.\*  
107: gb\_est30.\*  
108: gb\_est31.\*  
109: gb\_est32.\*  
110: gb\_est41.\*  
111: gb\_est42.\*  
112: gb\_est43.\*  
113: gb\_est44.\*  
114: gb\_est45.\*  
115: gb\_est46.\*  
116: gb\_est47.\*

117: gb\_est48:\*  
118: gb\_est49:\*  
119: gb\_est50:\*  
120: gb\_est51:\*  
121: gb\_est52:\*  
122: gb\_est53:\*  
123: gb\_est54:\*  
124: gb\_est55:\*  
125: gb\_est56:\*  
126: gb\_est57:\*  
127: gb\_est58:\*  
128: gb\_est59:\*  
129: gb\_est60:\*  
130: gb\_est61:\*  
131: gb\_est62:\*  
132: gb\_est63:\*  
133: gb\_est64:\*  
134: gb\_est65:\*  
135: gb\_est66:\*  
136: gb\_est67:\*  
137: gb\_est68:\*  
138: gb\_est69:\*  
139: gb\_est70:\*  
140: gb\_est71:\*  
141: gb\_est72:\*  
142: gb\_est73:\*  
143: gb\_est74:\*  
144: gb\_est75:\*  
145: gb\_est76:\*  
146: gb\_est77:\*  
147: gb\_est78:\*  
148: gb\_est79:\*  
149: gb\_est80:\*  
150: gb\_est81:\*  
151: gb\_est82:\*  
152: gb\_est83:\*  
153: gb\_est84:\*  
154: gb\_est85:\*  
155: gb\_est86:\*  
156: gb\_est87:\*  
157: gb\_est88:\*  
158: gb\_est89:\*  
159: gb\_est90:\*  
160: gb\_est91:\*  
161: gb\_est92:\*  
162: gb\_est93:\*  
163: gb\_est94:\*  
164: gb\_est95:\*  
165: gb\_est96:\*  
166: gb\_est97:\*  
167: gb\_est98:\*  
168: gb\_est99:\*  
169: gb\_est100:\*  
170: gb\_est101:\*  
171: gb\_est102:\*  
172: gb\_est103:\*  
173: gb\_est104:\*  
174: gb\_est105:\*  
175: gb\_est106:\*  
176: gb\_est107:\*  
177: gb\_est108:\*  
178: gb\_est109:\*  
179: gb\_est110:\*  
180: gb\_est111:\*  
181: gb\_est112:\*  
182: gb\_est113:\*  
183: gb\_est114:\*  
184: gb\_est115:\*  
185: gb\_est116:\*  
186: gb\_est117:\*  
187: gb\_est118:\*  
188: gb\_est119:\*  
189: gb\_est120:\*

190: em\_gss\_pln1:\*  
191: em\_gss\_pln2:\*  
192: em\_gss\_pro:\*  
193: em\_gss\_rod1:\*  
194: em\_gss\_rod2:\*  
195: em\_gss\_rod3:\*  
196: em\_gss\_rod4:\*  
197: em\_gss\_rod5:\*  
198: em\_gss\_vit1:\*  
199: em\_gss\_vit2:\*  
200: em\_gss\_vit3:\*  
201: gb\_gss1:\*  
202: gb\_gss2:\*  
203: gb\_gss3:\*  
204: gb\_gss4:\*  
205: gb\_gss5:\*  
206: gb\_gss6:\*  
207: gb\_gss7:\*  
208: gb\_gss8:\*  
209: gb\_gss9:\*  
210: gb\_gss10:\*  
211: gb\_gss11:\*  
212: gb\_gss12:\*  
213: gb\_gss13:\*  
214: gb\_gss14:\*  
215: gb\_gss15:\*  
216: gb\_gss16:\*  
217: gb\_gss17:\*  
218: gb\_gss18:\*  
219: gb\_gss19:\*  
220: gb\_gss20:\*  
221: gb\_gss21:\*  
222: gb\_gss22:\*  
223: gb\_gss23:\*  
224: gb\_gss24:\*  
225: gb\_gss25:\*  
226: gb\_gss26:\*  
227: gb\_gss27:\*  
228: gb\_gss28:\*  
229: gb\_gss29:\*  
230: gb\_gss30:\*  
231: gb\_gss31:\*  
232: gb\_gss32:\*  
233: gb\_gss33:\*  
234: gb\_gss34:\*

pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query %			DB ID	Description
		Match	Length	ID		
1	141	100.0	533	163	BE129089	BE129089 894020B03
2	43	30.5	500	113	AW256120	AW256120 687062E03
3	42.4	30.1	932	229	CNS00720	AL066742 Drosophila
4	41.4	29.4	466	117	AW566109	AW566109 660062B03
5	41.4	29.4	577	116	AW498392	AW498392 660047C10
6	39.8	28.2	508	104	AJ280717	AJ280717 4A3A-AAS-
7	39.8	28.2	1375	119	AW727483	AW727483 GA_Ea001
8	38.6	27.4	651	105	AL508953	AL508953 AL508953
9	37.6	26.7	430	102	AI815951	AI815951 au44f02.y
10	37.4	26.5	179	9	AA594531	AA594531 nl94h11.s
11	37.2	26.4	401	175	D22780	D22780 RICC1302A.R
12	37.2	26.4	925	229	CNS0091P	AL053013 Drosophila
13	37	26.2	464	31	AV620164	AV620164 AV620164
14	36.8	26.1	846	231	CNS04J01	AL293770 Tetraodon
15	36.8	26.1	945	230	CNS02D8L	AL192126 Tetraodon
16	36.6	26.0	513	105	AL450546	AL450546 AL450546
17	36.6	26.0	925	229	CNS0091P	AL053013 Drosophila
18	36.6	26.0	1039	230	CNS02V8F	AL215448 Tetraodon





TITLE Anopheles gambiae pilot gene discovery project: identification of mosquito innate immunity genes from expressed sequence tags

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (12), 6619-6624 (2000)

MEDLINE 20300950

COMMENT Contact: Dimopoulos G  
Fotis C. Kafatos laboratory  
European Molecular Biology Laboratory  
Meyerhofstrasse 1, 69117 Heidelberg, Germany.  
Location/Qualifiers  
I. .508  
/organism="Anopheles gambiae"  
/strain="4A r/r"  
/db\_xref="taxon:7165"  
/clone="4A3A-AAS-G-09"  
/clone.lib="Anopheles gambiae immune competent 4A3A"  
/cell\_line="immune competent 4A3A"  
/lab\_host="E. coli DH10B"  
/note=Vector: p773D-Pac (Pharmacia) with a modified polylinker; Site\_1: EcoRI; Site\_2: NotI; sequenced from forward priming site which reads from the 3' end of the cDNA. The 4A3A is a directionally cloned and normalized cDNA library that was constructed from the 4A3A cell line oligo-T primed cDNA according to: Bonaldo, Lennon & Soares (1996). Normalization and Subtraction: Two approaches To Facilitate Gene Discovery. Genome Research 6, 791-806.

BASE COUNT 58 a 168 c 175 g 107 t

ORIGIN

Query Match 28.2%; Score 39.8; DB 104; Length 508;  
Best Local Similarity 64.8%; Pred. No. 4.5;  
Matches 59; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 46 gtgacgtctgtccgccgtgagcgcctggcgaggagagcgcaagcgctcgacgggtcc 105  
| | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 146 GAGTCGAAGTGTCCCGCGGACAGGAGGAGGTGGGAAACACCTCGATCGTGCG 205  
| | | | | | | | | | | | | | | | | | | | | | | | | | | |  
QY 106 gtgcggcgcttcgctcagctgcaccgcgcga 136  
| | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 206 GTGC GGCGCAGCGGGAAGCTCGCGCGGGCA 236

RESULT 7

AW727483/c

LOCUS AW727483.1 1375 bp mRNA EST 16-NOV-2000

DEFINITION GA\_Ea0012118 Gossypium arboreum 7-10 dpa fiber library Gossypium arboreum CDNA clone GA\_Ea0012118, mRNA sequence.

ACCESSION AW727483

VERSION AW727483.1 GI:7625038

KEYWORDS EST.

SOURCE Gossypium arboreum.

ORGANISM Gossypium arboreum  
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Malvales; Malvaceae; Gossypium.  
1. (bases 1 to 1375)  
Wing, R.A., Frisch, D., Yu, Y., Main, D., Rambo, T., Simmons, J., Henry D., Wood, T.C., Leslie, A., and Wilkins, T.A.  
An integrated analysis of the genetics, development, and evolution of the cotton fiber  
Unpublished (2000)  
Contact: Wing RA  
Clemson University  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Tel.: 864 656 7288  
Fax: 864 656 4293  
Email: rwing@clemson.edu  
High quality sequence stop: 1375.  
Location/Qualifiers  
I. .1375

FEATURES source





was constructed by Dr. Claudio Schneider (LNCIB-Area  
Science Park, Trieste, Italy). "

45 a	180 c	151 g	54 t
------	-------	-------	------

BASE COUNT

BASE COUNT  
ORIGIN

[illegible]

RESULT 10  
AA594531  
LOCUS  
DEFINITION  
25-SEP-1997  
EST  
mRNA  
n194hl1.s1 NCI-CGAP.Col0 Homo sapiens cDNA clone IMAGE:1058373 3'  
similar to gb:M60484 ratl PROTEIN PHOSPHATASE PP2A-BETA, CATALYTIC  
SUBUNIT (HUMAN);, mRNA sequence.  
ACCESSION  
AA594531  
VERSION  
AA594531.1 GI:2409881  
KEYWORDS  
EST.  
SOURCE  
human.  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1 (bases 1 to 179)  
AUTHORS  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
TITLE  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP).  
JOURNAL  
Tumor Gene Index  
COMMENTS  
Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.  
Tel.: (301) 496-1550  
Email: [Robert.Strausberg@nih.gov](mailto:Robert.Strausberg@nih.gov)  
Tissue Procurement: Ilian Kirsch, M.D., Michael R. Emmert-Buck, M.D.,  
Ph.D.  
cDNA Library Preparation: M. Bento Soares, Ph.D.  
cDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
[www-bio.llnl.gov/bbrp/image/image.html](http://www-bio.llnl.gov/bbrp/image/image.html)  
Insert Length: 1071 Std Error: 0.00  
Seq primer: -40ml3 fwd. ET from Amersham.

BASE COUNT	20 a	74 c	68 g	17 t
ORIGIN	benito Soares and M. Fatima Ronaldo (N-Soares).			

Query Match	26.5%;	Score 37.4;	DB 9;	Length 179;
Best Local Similarity	55.9%;	Pred. No. 17;		
Matches 71;	Conservative 0;	Mismatches 56;	Indels 0;	Gaps
Qy	7	gaagccctgggctctcttgactaccgcggtggagcgcgtgacgtgtcgtaccgcgtg	66	
Db	40	GACCCCGCCGCCGCCCTTCGCCCGCCGCCCGCGCGCGGAGTCGGTGAAGACACGGGTG	99	
Qy	67	agcgcgtgctgggagagcgaagggctctcgacgggttcctgycggggcttcggtcagctg	126	
Db	100	AGTGTCCGGGGACGCGCGCGGGTCTCTCGGCTAGCTCTCGCGGCACGAGCGCCGCCG	159	
Qy	127	cacccgc	133	
Db	160	CTGCGC	166	

RESULT 11

D22780/c	401 bp	mRNA	EST	08-JUL-1999
LOCUS				
DEFINITION	R1C1302A rice callus Oryza sativa subsp. japonica cDNA clone			
	C1302_1A, mRNA sequence.			
ACCESSION	D22780			
VERSION	D22780.1	GI:426710		
KEYWORDS	EST.			
SOURCE	Oryza sativa subsp. japonica.			
ORGANISM	Oryza sativa subsp. japonica			
	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;			
	Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae;			
	Oryza.			
REFERENCE	1 (bases 1 to 401)			
AUTHORS	Sasaki, T. and Minobe, Y.			
TITLE	Rice cDNA from callus			
JOURNAL	Unpublished (1994)			
COMMENT	Contact: Takuji Sasaki			
	National Institute of Agrobiological Resources			
	Rice Genome Research Program			
	2-1-2 Kannondai, Tsukuba			
	Ibaraki,			
	Japan 305			
	Tel: 0298-38-7441			
	Fax: 0298-38-7468			
	Email: tsasaki@agr.affrc.go.jp			

```

BASE COUNT      83 a    94 c   165 g    52 t      7 others
ORIGIN
      1  ggcgcgagccctgggctccttgcattaccgccggtggcgccgtgacgctgtcgtac 50
      2  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
      3  259  GC CGCGTCA CCTG CCGCTTCTCCACGACCGCGCGGGTGCCCTGCTGCCGCGCGG 200
      4  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
      5  QY   61  csgctgagccctgcggaggagcgcaagcctcgacggttcggtccgggcttcggt 120
      6  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
      7  Db   199  CAGCTCTCGCGCGCGGCGNCAGACGTGCAGTGTGCGCGACTGGCGCTTGCTTCA 140
      8  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
      9  QY   121  cagctgcaccgcgcgcgcgag 141

```







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 2, 2001, 00:04:52 ; Search time 128.09 Seconds  
(without alignments)  
642.620 Million cell updates/sec

Title: US-09-331-723a-4

Perfect score: 141

Sequence: 1 gccgcagccctggctc.....agctgcacccgcgcagcag 141

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 678276 seqs, 291890651 residues

Total number of hits satisfying chosen parameters: 1356552

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N\_Geneseq\_0401.\*  
1: /SID56/gcgdata/geneseq/geneseq/NA1980.DAT.\*  
2: /SID56/gcgdata/geneseq/geneseq/NA1981.DAT.\*  
3: /SID56/gcgdata/geneseq/geneseq/NA1982.DAT.\*  
4: /SID56/gcgdata/geneseq/geneseq/NA1983.DAT.\*  
5: /SID56/gcgdata/geneseq/geneseq/NA1984.DAT.\*  
6: /SID56/gcgdata/geneseq/geneseq/NA1985.DAT.\*  
7: /SID56/gcgdata/geneseq/geneseq/NA1986.DAT.\*  
8: /SID56/gcgdata/geneseq/geneseq/NA1987.DAT.\*  
9: /SID56/gcgdata/geneseq/geneseq/NA1988.DAT.\*  
10: /SID56/gcgdata/geneseq/geneseq/NA1989.DAT.\*  
11: /SID56/gcgdata/geneseq/geneseq/NA1990.DAT.\*  
12: /SID56/gcgdata/geneseq/geneseq/NA1991.DAT.\*  
13: /SID56/gcgdata/geneseq/geneseq/NA1992.DAT.\*  
14: /SID56/gcgdata/geneseq/geneseq/NA1993.DAT.\*  
15: /SID56/gcgdata/geneseq/geneseq/NA1994.DAT.\*  
16: /SID56/gcgdata/geneseq/geneseq/NA1995.DAT.\*  
17: /SID56/gcgdata/geneseq/geneseq/NA1996.DAT.\*  
18: /SID56/gcgdata/geneseq/geneseq/NA1997.DAT.\*  
19: /SID56/gcgdata/geneseq/geneseq/NA1998.DAT.\*  
20: /SID56/gcgdata/geneseq/geneseq/NA1999.DAT.\*  
21: /SID56/gcgdata/geneseq/geneseq/NA2000.DAT.\*  
22: /SID56/gcgdata/geneseq/geneseq/NA2001.DAT.\*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	% Query Match	Length	DB ID	Description
1	141	100.0	141	19	V46291
2	141	100.0	1838	20	232119
3	141	100.0	1838	21	245282
4	139.4	98.9	2573	19	V46297
5	139.4	98.9	3381	19	V46298
6	136.2	96.6	3381	18	T78597
7	136.2	96.6	3383	18	T60965
8	38.8	27.5	2900	13	Q20372
9	38.2	27.1	30001	18	T61016
10	38.2	27.1	30001	20	X05110
11	38	27.0	2310	20	221502

C	12	35.8	25.4	1288	20	V84525
	13	35.4	25.1	3035	21	A63462
	14	35.4	25.1	42000	21	A63349
	15	35.4	25.1	63164	21	A63348
C	16	34.6	24.5	1257	21	A75110
	17	34.6	24.5	1286	21	A26456
C	18	34.6	24.5	1342	21	A26424
	19	34.2	24.3	2657	20	X84348
	20	34.2	24.3	3510	21	C65299
	21	34.2	24.3	43280	18	T80413
C	22	34	24.1	1215	18	T73246
	23	34	24.1	1329	21	A4984
C	24	34	24.1	2885	20	X24071
	25	33.8	24.0	1874	20	X28279
	26	33.8	24.0	1874	20	X28278
	27	33.8	24.0	8937	13	Q20502
	28	33.8	24.0	10706	18	T46941
C	29	33.6	23.8	24379	18	T93095
	30	33.6	23.8	24379	19	V25925
C	31	33.4	23.7	738	20	X86968
	32	33.4	23.7	2490	18	T91358
	33	33.4	23.7	2634	19	V23480
	34	33.4	23.7	2634	21	A13906
	35	33.4	23.7	6436	20	X86965
	36	33.4	23.7	17612	19	V23494
	37	33.4	23.7	17612	21	A13905
	38	33.4	23.7	58857	21	A58471
	39	33.2	23.5	780	18	T76882
	40	33.2	23.5	900	18	T76877
C	41	33.2	23.5	1215	19	V23483
C	42	33.2	23.5	1215	21	A13894
	43	33.2	23.5	5516	18	T76880
C	44	33.2	23.5	17612	19	V23494
C	45	33.2	23.5	17612	21	A13905

#### ALIGNMENTS

RESULT 1  
V46291  
ID V46291 standard; DNA; 141 BP.  
XX  
AC V46291;  
XX  
DT 16-OCT-1998 (first entry)  
XX  
DE C. reinhardtii porphyrin herbicide resistance domain genomic DNA.  
XX  
KW Herbicide resistance; porphyrin; protoporphyrinogen oxidase; PPO;  
XX crop plant; selective weed control; ss.  
XX  
OS Chlamydomonas reinhardtii.  
XX  
PN WO9829554-A1.  
XX  
PD 09-JUL-1998.  
XX  
PF 27-DEC-1996; 96WO-US20415.  
XX  
PR 27-DEC-1996; 96WO-US20415.  
XX  
PA (SUMO) SUMITOMO CHEM CO LTD.  
XX (UYDU-) UNIV DUKE.  
XX Boynton JE, Gillham NW, Ishige F, Randolph-Anderson BL;  
XX Sato R;  
XX WPI: 1998-388134/33.  
XX P-PSDB; W64450.  
XX Conferring resistance to herbicides that inhibit proto-porphyrinogen  
XX oxidase - by introducing DNA that encodes a resistant form of the

Human secreted pro  
Streptomyces globi  
Streptomyces globi  
cDNA encoding a hu  
Human secreted pro  
Human secreted pro  
Stealth virus nucl  
Maize Scarecrow nu  
Tylactone synthase  
Delta-p62 coding s  
DNA encoding human  
Human p62 DNA. HO  
N. tabacum protopo  
N. tabacum protopo  
Nf1 gene. Homo sa  
Human neurofibroma  
Streptomyces feno  
Streptomyces roseo  
PHA acetacetyl-Co  
Orf virus strain N  
Pseudomonas orfiv s  
Pseudomonas alcali  
DNA containing PHA  
Pseudomonas Xpc, O  
Nucleotide sequenc  
Pseudomonas alcali  
Parapox virus Hind  
Parapox virus Hind  
Pseudomonas Xcps s  
Pseudomonas alcali  
Parapox virus Hind  
Pseudomonas Xpc, O  
Pseudomonas alcali

PT enzyme, also resistant plants and algae, mutant DNA, vectors and  
PT transformed microorganisms  
PS Example 10; Page 67; 108pp; English.  
XX  
CC This genomic DNA sequence encodes a fragment of the Chlamydomonas  
CC -reinhardtii porphyrin herbicide resistance domain which is used in a  
CC method to investigate resistance to herbicides that inhibit  
CC protoporphyrinogen oxidase (PPO). The method is applied to crop plants,  
CC allowing use of herbicides for selective weed control. The method can be  
CC used to identify compounds that are, or are not, PPO-inhibitors.  
XX  
SQ Sequence 141 BP; 15 A; 53 C; 54 G; 19 T; 0 other;  
  
Query Match 100.0%; Score 141; DB 19; Length 141;  
Best Local Similarity 100.0%; Pred. No. 2e-22;  
Matches 141; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 gccgcgagccctggctcttcgactaccgcgcggtggcgccgtgacgtgtgtac 60  
|||||  
Db 1 gccgcgagccctggctcttcgactaccgcgcggtggcgccgtgacgtgtgtac 60  
|||||  
Qy 61 ccgctgagcgcgtgcccggagagcgcagggcctgcgcggttcggtccggttcggt 120  
|||||  
Db 61 ccgctgagcgcgtgcccggagagcgcagggcctgcgcggttcggtccggttcggt 120  
|||||  
Qy 121 cagctgcacccgcgcgcag 141  
|||||  
Db 121 cagctgcacccgcgcgcag 141  
|||||  
RESULT 2  
232119  
ID 232119 standard; cDNA; 1838 BP.  
XX  
AC 232119;  
XX  
DT 12-JAN-2000 (first entry)  
XX  
DE Chlamydomonas reinhardtii protoporphyrinogen oxidase encoding cDNA.  
XX  
KW Protoporphyrinogen oxidase; PPO; herbicide; inhibition; resistance;  
KW malignant cell growth; phototropic herbicide; ss.  
XX  
OS Chlamydomonas reinhardtii.  
XX  
FH Key Location/Qualifiers  
FT CDS 2..1793  
FT /tag= a  
FT /product= "protoporphyrinogen oxidase"  
XX  
PN EP955380-A2.  
XX  
PD 10-NOV-1999.  
XX  
PF 09-APR-1999; 99EP-0107037.  
XX  
PR 10-APR-1998; 98JP-0099619.  
XX  
PA (SUMO ) SUMITOMO CHEM CO LTD.  
XX  
PI Shimokawatoko Y, Nishio S;  
XX  
FI WPI; 1999-612726/53.  
XX  
DR P-PSDB; Y49538.  
XX  
PT Evaluating ability of compounds to inhibit protoporphyrinogen oxidase  
PT activity for development of herbicides -  
XX  
PS Claim 22; Page 28-31; 46pp; English.  
XX  
CC A method has been developed for evaluating the ability of a compound to

CC inhibit protoporphyrinogen oxidase (PPO) activity. The method comprises:  
CC (a) culturing a transformant expressing a protoporphyrinogen oxidase  
CC gene in the presence and absence of a test compound to measure a growth  
CC rate; and (b) determining the ability of the compound to inhibit PPO  
CC activity by comparing growth rates. The method is useful for the  
CC development of PPO inhibiting-type herbicides. Also the method is useful  
CC for screening medications having the ability for inhibiting the  
CC malignant cell growth in addition to evaluating the potency of a  
CC phototropic herbicide, screening a compound having a herbicidal  
CC activity. The PPO gene can be used to obtain or generate a gene  
CC resistance to a PPO activity inhibiting agent e.g. a phototropic  
CC herbicide. The present sequence encodes Chlamydomonas reinhardtii-  
XX derived PPO.  
XX  
SQ Sequence 1838 BP; 302 A; 619 C; 634 G; 283 T; 0 other;  
  
Query Match 100.0%; Score 141; DB 20; Length 1838;  
Best Local Similarity 100.0%; Pred. No. 1.6e-22;  
Matches 141; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 gccgcgagccctggctcttcgactaccgcgcggtggcgccgtgacgtgtgtac 60  
|||||  
Db 1130 gccgcgagccctggctcttcgactaccgcgcggtggcgccgtgacgtgtgtac 1189  
|||||  
Qy 61 ccgctgagcgcgtgcccggagagcgcagggcctgcgcggttcggtccggttcggt 120  
|||||  
Db 1190 ccgctgagcgcgtgcccggagagcgcagggcctgcgcggttcggtccggttcggt 1249  
|||||  
Qy 121 cagctgcacccgcgcgcag 141  
|||||  
Db 1250 cagctgcacccgcgcgcag 1270  
|||||  
RESULT 3  
245282  
ID 245282 standard; DNA; 1838 BP.  
XX  
AC 245282;  
XX  
DT 27-MAR-2000 (first entry)  
XX  
DE DNA encoding an algal protoporphyrin IX oxidase protein.  
XX  
KW Weed control; weed activity; plant; protoporphyrin IX oxidase; PPO;  
KW herbicide; chlormethoxynil; acifluorfen; protoporphyrin X; ss.  
XX  
OS Chlamydomonas reinhardtii.  
XX  
FH Key Location/Qualifiers  
FT CDS 2..1693  
FT /tag= a  
FT /product= "protoporphyrin IX oxidase"  
XX  
PN EP953646-A2.  
XX  
PD 03-NOV-1999.  
XX  
PF 30-APR-1999; 99EP-0108463.  
XX  
PR 30-APR-1998; 98JP-0120553.  
PR 02-OCT-1998; 98JP-0281127.  
PR 20-NOV-1998; 98JP-0330981.  
PR 02-MAR-1999; 99JP-0054730.  
XX  
PA (SUMO ) SUMITOMO CHEM CO LTD.  
XX  
PI Nakajima H, Nagasawa A;  
XX  
FI WPI; 2000-088762/08.  
XX  
DR P-PSDB; Y54101.  
XX  
PT Method of making plants resistant to weed control compounds by

PT introducing gene encoding protein into plant cell -  
 XX Example 15; Page 72-76; 119pp; English.

XX The present sequence encodes an algal protoporphyrin IX oxidase (PPO).  
 CC The sequence was used in the course of the invention. The specification  
 CC describes a method of making plants resistant to weed control compounds.  
 CC The method comprises introducing a gene encoding a protein into a plant  
 CC cell, where the protein has a specific affinity for a substance which  
 CC is concerned with the weed activity of a weed control compound, and  
 CC e.g. protoporphyrin IX; has no ability to modify this protein; and  
 CC is virtually free from framework regions in an immunoglobulin; and  
 CC expressing the gene. The gene of the invention is used for producing a  
 CC plant resistant to specified compounds, such as protoporphyrin IX  
 CC oxidase (PPO) inhibitory-type herbicides, e.g. chlormethoxynil,  
 CC acifluorfen, etc.. When PPO is inhibited, protoporphyrin IX (the  
 CC substrate of PPO) accumulates in the plant cells, and is metabolised to  
 CC form protoporphyrin X, followed by formation of active oxygen in the  
 CC presence of protoporphyrin X and light, which damages cell functions.

XX Sequence 1838 BP; 302 A; 619 C; 534 G; 283 T; 0 other;

Query Match 100.0%; Score 141; DB 21; Length 1838;  
 Best Local Similarity 100.0%; Pred. No. 1.6e-22;  
 Matches 141; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 gccgcgagccctgggtcctctgactaccgccggtggcgccgtgacgtctgtac 60  
 Db |||||||||||||||||||||||||||||||||||||||||||||||||||||  
 1130 gccgcgagccctgggtcctctgactaccgccggtggcgccgtgacgtctgtac 1189  
 QY 61 ccgctgagccctggcgagagcgcaagcctcgagcgggtccggtccgggttcggt 120  
 Db |||||||||||||||||||||||||||||||||||||||||||||||||||||  
 1190 ccgctgagccctggcgagagcgcaagcctcgagcgggtccggtccgggttcggt 1249  
 QY 121 cagctgacccgcgcgcag 141  
 Db |||||||||||||||||||||||||||||||||||||||||||||||||||||  
 1250 cagctgacccgcgcgcag 1270

RESULT 4  
 V46297 ID V46297 standard; DNA; 2573 BP.  
 AC V46297;  
 XX  
 XX 16-OCT-1998 (first entry)  
 DT  
 XX C. reinhardtii protoporphyrinogen oxidase DNA fragment Xho/Pmac2.6.  
 DE  
 XX Herbicide resistance; porphyrin; protoporphyrinogen oxidase; PPO;  
 KW crop plant; selective weed control; ss.  
 XX Chlamydomonas reinhardtii.  
 OS  
 PN W09829554-A1.  
 XX  
 XX 09-JUL-1998.  
 PD  
 XX 27-DEC-1996; 96WO-US20415.  
 PF  
 XX 27-DEC-1996; 96WO-US20415.  
 PR  
 XX (SUMO ) SUMITOMO CHEM CO LTD.  
 PA (UYDU-) UNIV DUKE.  
 XX  
 XX Boynton JE, Gillham NW, Ishige F, Randolph-Anderson BL;  
 PI -Sato R;  
 XX WPI; 1998-388134/33.  
 DR  
 XX Conferring resistance to herbicides that inhibit proto-porphyrinogen  
 XX oxidase - by introducing DNA that encodes a resistant form of the

PT introducing gene encoding protein into plant cell -  
 XX Example 15; Page 72-76; 119pp; English.

XX The present sequence encodes an algal protoporphyrin IX oxidase (PPO).  
 CC The sequence was used in the course of the invention. The specification  
 CC describes a method of making plants resistant to weed control compounds.  
 CC The method comprises introducing a gene encoding a protein into a plant  
 CC cell, where the protein has a specific affinity for a substance which  
 CC is concerned with the weed activity of a weed control compound, and  
 CC e.g. protoporphyrin IX; has no ability to modify this protein; and  
 CC is virtually free from framework regions in an immunoglobulin; and  
 CC expressing the gene. The gene of the invention is used for producing a  
 CC plant resistant to specified compounds, such as protoporphyrin IX  
 CC oxidase (PPO) inhibitory-type herbicides, e.g. chlormethoxynil,  
 CC acifluorfen, etc.. When PPO is inhibited, protoporphyrin IX (the  
 CC substrate of PPO) accumulates in the plant cells, and is metabolised to  
 CC form protoporphyrin X, followed by formation of active oxygen in the  
 CC presence of protoporphyrin X and light, which damages cell functions.

PT enzyme, also resistant plants and algae, mutant DNA, vectors and  
 PT transformed microorganisms

PS Example 7; Page 70-71; 108pp; English.

XX This genomic DNA sequence encodes protoporphyrinogen oxidase from  
 CC Chlamydomonas reinhardtii (fragment Xho/Pmac2.6). This sequence is used  
 CC in a method which investigates resistance to herbicides that inhibit  
 CC protoporphyrinogen oxidase (PPO) using a porphyrin herbicide resistance  
 CC domain. The method is applied to crop plants, allowing use of herbicides  
 CC for selective weed control. The method can be used to identify compounds  
 CC that are, or are not, PPO-inhibitors.

XX Sequence 2573 BP; 470 A; 859 C; 735 G; 509 T; 0 other;

Query Match 98.9%; Score 139.4; DB 19; Length 2573;  
 Best Local Similarity 99.3%; Pred. No. 3.4e-22;  
 Matches 140; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 gccgcgagccctgggtcctctgactaccgccggtggcgccgtgacgtctgtac 60  
 Db |||||||||||||||||||||||||||||||||||||||||||||||||||||  
 255 gccgcgagccctgggtcctctgactaccgccggtggcgccgtgacgtctgtac 314  
 QY 61 ccgctgagccctggcgagagcgcaagcctcgagcgggtccggtccgggttcggt 120  
 Db |||||||||||||||||||||||||||||||||||||||||||||||||||||  
 315 ccgctgagccctggcgagagcgcaagcctcgagcgggtccggtccgggttcggt 374  
 QY 121 cagctgacccgcgcgcag 141  
 Db |||||||||||||||||||||||||||||||||||||||||||||||||||||  
 375 cagctgacccgcgcgcag 395

RESULT 5  
 V46298 ID V46298 standard; DNA; 3381 BP.  
 AC V46298;  
 XX  
 XX 16-OCT-1998 (first entry)  
 DT  
 XX C. reinhardtii protoporphyrinogen oxidase DNA fragment Xho3.4.  
 DE  
 XX Herbicide resistance; porphyrin; protoporphyrinogen oxidase; PPO;  
 KW crop plant; selective weed control; ss.  
 XX Chlamydomonas reinhardtii.  
 OS  
 PN W09829554-A1.  
 XX  
 XX 09-JUL-1998.  
 PD  
 XX 27-DEC-1996; 96WO-US20415.  
 PF  
 XX 27-DEC-1996; 96WO-US20415.  
 PR  
 XX (SUMO ) SUMITOMO CHEM CO LTD.  
 PA (UYDU-) UNIV DUKE.  
 XX  
 XX Boynton JE, Gillham NW, Ishige F, Randolph-Anderson BL;  
 PI Sato R;  
 XX WPI; 1998-388134/33.  
 DR  
 XX Conferring resistance to herbicides that inhibit proto-porphyrinogen  
 XX oxidase - by introducing DNA that encodes a resistant form of the  
 PT enzyme, also resistant plants and algae, mutant DNA, vectors and  
 PT transformed microorganisms

XX Example 7; Page 83-85; 108pp; English.

XX This genomic DNA sequence encodes protoporphyrinogen oxidase from  
 CC Chlamydomonas reinhardtii (fragment Xho3.4). This sequence is used

CC in a method which investigates resistance to herbicides that inhibit  
CC protoporphyrinogen oxidase (PPO) using a porphyrin herbicide resistance  
CC domain. The method is applied to crop plants, allowing use of herbicides  
CC for selective weed control. The method can be used to identify compounds  
CC that are, or are not, PPO-inhibitors.

XX SQ Sequence 3381 BP; 544 A; 1058 C; 1010 G; 669 T; 0 other;

Query Match 98.9%; Score 139.4; DB 19; Length 3381;  
Best Local Similarity 99.3%; Pred. No. 3.3e-22;  
Matches 140; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 gccgcagagccctggctcttcgactaccgcgcggtggcgccgctgacgtcgctac 60  
|||||  
Db 255 gccgcagagccctggctcttcgactaccgcgcggtggcgccgctgacgtcgctac 314  
|||||  
Qy 61 ccgctgagcgcgtgagcgagcgcaaggcctcgacgaggtcgctgcccggctcggt 120  
|||||  
Db 315 ccgctgagcgcgtgagcgagcgcaaggcctcgacgaggtcgctgcccggctcggt 374  
|||||  
Qy 121 cagctgacccgcgcacgacg 141  
|||||  
Db 375 cagctgacccgcgcacgacg 395  
|||||

## RESULT 6

T78597/c  
ID T78597 standard; DNA; 3381 BP.

XX AC T78597;

XX DT 03-NOV-1997 (first entry)

XX DE Porphyrin-accumulating type herbicide resistant DNA fragment.

XX KW Plant; algal; transformation; porphyrin herbicide; weed control;  
KW oxadiazon; oxyfluorfen; ss.

XX OS Chlamydomonas reinhardtii - strain RS-3.

XX PN W09704089-A2.

XX PD 06-FEB-1997.

XX PF 19-JUL-1996; 96WO-US11999.

XX PR 20-JUL-1995; 95WO-US09098.

XX PA (SUMO ) SUMITOMO CHEM CO LTD.  
PA (UYDU-) UNIV DUKE.

XX PI Boynton J, Gillham NW, Harris EH, Sato R;

XX DR WPI; 1997-132630/12.

XX PT Chlamydomonas reinhardtii derived DNA conferring resistance to  
PT porphyrin herbicides, used to transform plant or algal cells to  
PT confer selective herbicide resistance

XX PS Claim 14; Page 40-41; 50pp; English.

XX CC The present sequence represents an isolated, purified DNA fragment  
CC that confers resistance to porphyrin-accumulating type herbicides  
CC when expressed in plant or algal cells. The DNA fragment can be used  
CC in plants and algae to confer resistance to herbicides such as  
CC N-(4-chloro-2-fluoro-5-propargyloxy) phenyl-3,4,5,6  
CC tetrahydrothialimide, oxadiazon and oxyfluorfen. It permits the use  
CC of porphyrin herbicides during crop cultivation and thus facilitates  
CC weed control in crops. The gene can also be used as a resistance-type  
CC genetic marker in plant genetic engineering.

XX SQ Sequence 3381 BP; 670 A; 1008 C; 1058 G; 645 T; 0 other;

Query Match 96.6%; Score 136.2; DB 18; Length 3381;  
Best Local Similarity 97.9%; Pred. No. 1.6e-21;  
Matches 138; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 gccgcagagccctggctcttcgactaccgcgcggtggcgccgctgacgtcgctac 60  
|||||  
Db 3127 gccgcagagccctggctcttcgactaccgcgcggtggcgccgctgacgtcgctac 3068  
|||||  
Qy 61 ccgctgagcgcgtgagcgagcgcaaggcctcgacgaggtcgctgcccggctcggt 120  
|||||  
Db 3067 ccgctgagcgcgtgagcgagcgcaaggcctcgacgaggtcgctgcccggctcggt 3008  
|||||  
Qy 121 cagctgacccgcgcacgacg 141  
|||||  
Db 3007 cagctgacccgcgcacgacg 2987  
|||||

## RESULT 7

T60965/c  
ID T60965 standard; DNA; 3383 BP.

XX AC T60965;

XX DT 14-MAY-1997 (first entry)

XX DE Porphyrin-accumulating type herbicide resistance gene.

XX KW Herbicide resistance; porphyrin; transgenic plant; alga;  
KW weed control; selectable marker; ds.

XX OS Chlamydomonas reinhardtii strain RS-3.

XX PN W09704088-A1.

XX PD 06-FEB-1997.

XX PF 20-JUL-1995; 95WO-US09098.

XX PR 20-JUL-1995; 95WO-US09098.

XX PA (SUMO ) SUMITOMO CHEM CO LTD.

XX PA (UYDU-) UNIV DUKE.

XX PI Boynton J, Gillham NW, Harris EH, Sato R;

XX DR WPI; 1997-132629/12.

XX PT DNA conferring resistance to porphyrin herbicides - isolated from  
PT Chlamydomonas reinhardtii, useful to transform plant or algal cells  
PT to confer selective herbicide resistance

XX PS Claim 14; Page 39-41; 52pp; English.

XX CC A porphyrin accumulating-type herbicide resistance gene (T60965)  
CC was isolated from Chlamydomonas reinhardtii strain RS-3. Sau3AI  
CC fragments of genomic DNA were ligated into SuperCos-1 vector and  
CC used to transform E. coli. The DNA was extracted and used to  
CC transform the normally sensitive strain CC48 to provide herbicide  
CC resistance to N-(4-chloro-2-fluoro-5-propargyloxy)phenyl-3,4,5,6-  
CC tetrahydrothialimide. The DNA was then analysed to provide the  
CC porphyrin herbicide resistance gene. The gene can be incorporated  
CC into a plasmid and used to confer resistance to porphyrin  
CC accumulating-type herbicides in plant or algal cells. It can also  
CC be used as a genetic resistance marker.

XX SQ Sequence 3383 BP; 671 A; 1008 C; 1057 G; 647 T; 0 other;

Query Match 96.6%; Score 136.2; DB 18; Length 3383;  
Best Local Similarity 97.9%; Pred. No. 1.6e-21;  
Matches 138; Conservative 0; Mismatches 3; Indels 0; Gaps 0;



```
Qy 1 gccgcgagccctgggtctcttcgactaccgcgcggtggcgccgtgacgctgtcgtac 60
    |||||
Db 3128 GCCGCCGAGCCCTGGGCTCTTCGACTACCCGCCGATGGCGCGCTGACGCTGCTGATAC 3069
    |||||
Qy 61 ccgctgagccgctgctgggagggcgcaagagcctcggacggttcgctgcccggcttcggt 120
    |||||
Db 3068 CCgCTgAGGCGCGTgCGGgAGGAGCGCAAGACTCGGACGGGTCCGTGCCGGGCTTCGCT 3009
    |||||
Qy 121 cagctgaccccgccgcaqcag 141
    |||||
Db 3008 CATCTGCACCGCGCAGCAG 2988

RESULT 8
Q20372
ID Q20372 standard; DNA; 2900 BP.
XX AC Q20372;
XX DT 08-APR-1992 (first entry)
XX DE P. glumae PGI lipase A gene and stabilisation/translocation function.
XX KW lipase-specific stabilisation/translocation protein; detergent;
XX KW cleaning system; ORF2; ss.
XX OS Pseudomonas glumae.
XX FH Location/Qualifiers
XX CDS 483..1559
    /tag= a
    /product= lipase_A
    /start= 1559..2620
    /tag= b
    /label= ORF2
    /note= "different reading frame to lipA gene"
XX EP464922-A.
XX 08-JAN-1992.
XX 26-JUN-1991; 91EP-0201627.
XX 17-OCT-1990; 90EP-0202772.
XX 06-JUL-1990; 90EP-0307440.
XX 09-JUL-1990; 90WO-EP01138.
XX (UNIL ) UNILEVER NV.
XX Bos JW, Frenken LCJ, Verrips CT, Visser C;
XX WPI: 1992-010138/02.
XX P-PSDB; R20177, R20178.
XX Microorganism for use in detergents - contains lipase gene and a
XX translocation or stabilisation gene, used to produce an efficient
XX stable lipase
XX Claim 16; Fig 2; 46pp; English.
XX A P. glumae PGI genomic library was constructed in cosmid vector c2RB.
XX -The lipA gene was isolated by screening with probes based on the
XX N-terminal amino acid sequence of the P. glumae lipase. Immediately
XX downstream of the lipA gene was found a second open reading frame,
XX designated ORF2. It is ORF2 which encodes the lipase-specific
XX stabilisation/translocation activity. See also Q20373 and Q20464.
XX
XX Sequence 2900 BP; 412 A; 1041 C; 1037 G; 410 T; 0 other;

Query Match 27.5%; Score 38.8; DB 13; Length 2900;
Best Local Similarity 55.1%; Pred. No. 1.2;
```

```
Matches 76; Conservative 0; Mismatches 62; Indels 0; Gaps 0;
Qy 2 ccgcgagccctgggtctcttcgactaccgcgcggtggcgccgtgacgctgtcgtacc 61
    |||||
Db 2046 ccgcgagccgctgggtctcttcgactaccgcgcggtggcgccgtgacgctgtcgtacc 2105
    |||||
Qy 62 cgtgagccgctgctgggagggcgcaagagcctcggacggttcgctgcccggcttcggtc 121
    |||||
Db 2106 atcagcgcgcgctggcgccgacgcacgctcgcgagtgaggcgccgcttcttcggcg 2165
    |||||
Qy 122 agctgaccccgccgcaqcgc 139
    |||||
Db 2166 acgagcagcgccgcaqcgc 2183

RESULT 9
T61016
ID T61016 standard; DNA; 30001 BP.
XX AC T61016;
XX DT 21-APR-1997 (first entry)
XX DE Total DNA sequence from cosmid clones LP(2)127 and LP(2)128.
XX KW Cosmid clone; LP(2)127; LP(2)128; chlortetracycline;
XX KW biosynthetic pathway; recombinant; production; antibiotic;
XX KW heterologous host; Streptomyces lividans; ss.
XX OS Streptomyces aureofaciens.
XX PN US5589385-A.
XX PD 31-DEC-1996.
XX PF 26-JUL-1990; 90US-0558039.
XX PR 22-SEP-1993; 93US-0125468.
XX PR 26-JUL-1990; 90US-0558039.
XX PR 26-JUL-1990; 90US-0558040.
XX PR 15-JAN-1992; 92US-0821109.
XX PR 15-JAN-1992; 92US-0821419.
XX (AMCY ) AMERICAN CYANAMID CO.
XX Fantini SE, Lotvin JA, Ryan MJ, Strathy N;
XX WPI: 1997-076853/07.
XX DNA encoding tetracycline biosynthetic pathway proteins -
XX specifically from Streptomyces aureofaciens for expression in
XX heterologous hosts, specifically S. lividans
XX Claim 3; Columns 17-44; 39pp; English.
XX The present sequence is the total DNA sequence from cosmid clones
XX LP(2)127 and LP(2)128, which contains a Streptomyces aureofaciens
XX (ATCC 13899) DNA that encodes the proteins of the entire
XX chlortetracycline biosynthetic pathway. The biosynthetic gene,
XX which can be expressed in heterologous hosts, especially
XX S. lividans, may be useful in the production of antibiotics.
XX
XX Sequence 30001 BP; 4055 A; 10939 C; 11019 G; 3988 T; 0 other;

Query Match 27.1%; Score 38.2; DB 18; Length 30001;
Best Local Similarity 54.7%; Pred. No. 1.3;
Matches 76; Conservative 0; Mismatches 63; Indels 0; Gaps 0;
Qy 3 ccgcgagccgctgggtctcttcgactaccgcgcggtggcgccgtgacgctgtcgtacc 62
    |||||
Db 14962 cgacgtcgctacggcgccgagccaccctctgctgcgcgcgagatgcggagccgga 15021
```



CC arachnids, nematodes and insects. Biosynthesis of spinosyns occurs via  
CC stepwise condensation and modification of carboxylic acid precursors  
CC generating a linear polyketide which is modified further. The DNA  
CC sequence 221501 contains 24 genes and open reading frames involved in  
CC spinosyn biosynthesis, Y39297-Y39319. The genes identified in S. spinosa  
CC as having a role in the production of spinosyns are useful to improve  
CC yields of spinosyns, and for creating new spinosyns e.g. by mutagenesis,  
CC or interruption of steps in spinosyn biosynthesis. The modified spinosyns  
CC may be a new insect control agent or serve as substrates for further  
CC chemical modification and the creation of new semi-synthetic spinosyns.  
CC The genes are also useful to isolate similar sequences from S. spinosa or  
CC other species by hybridization.  
XX  
SQ Sequence 2310 BP; 356 A; 759 C; 853 G; 342 T; 0 other;  
  
Query Match 27.0%; Score 38; DB 20; Length 2310;  
Best Local Similarity 57.6%; Pred. No. 1.8;  
Matches 68; Conservative 0; Mismatches 50; Indels 0; Gaps 0;  
  
Oy 1 gccgcgagccctggctctctgactaccgcgcgtggcgccgtgacgtgtgtgtac 60  
Dy 1213 gtcgcgcgagcgttgggtctcttcgcggagcgcgaagcgcgcgtgacgcgtg 1272  
  
Oy 61 ccgctgagcgcgtgcggagagcgcaagcgcgcgtgcgcgcgtgcgcgcgttcg 118  
Dy 1273 gtgatacgcgcgcgtgtacacgcgcgtggagcgcgcgtgcgcgcgtgcgcgcgcg 1330  
  
RESULT 12  
V84525/c  
ID V84525 standard; DNA; 1288 BP.  
XX  
AC V84525;  
XX  
DT 01-MAR-1999 (first entry)  
XX  
DE Human secreted protein gene 115 clone HDTA95.  
XX  
KW Human; secreted protein; fusion protein; gene therapy; protein therapy;  
KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;  
KW developmental abnormality; foetal deficiency; blood; allergy; renal; ds;  
KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;  
KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;  
KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;  
KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;  
KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.  
XX  
OS Homo sapiens.  
XX  
PN WO9854963-A2.  
XX  
PD 10-DEC-1998.  
XX  
PE 04-JUN-1998; 98WO-US11422.  
XX  
PR 18-DEC-1997; 97US-0070923.  
PR 06-JUN-1997; 97US-0048877.  
PR 06-JUN-1997; 97US-0048881.  
PR 06-JUN-1997; 97US-0048884.  
PR 06-JUN-1997; 97US-0048893.  
PR 06-JUN-1997; 97US-0048896.  
PR 06-JUN-1997; 97US-0048899.  
PR 06-JUN-1997; 97US-0048915.  
PR 06-JUN-1997; 97US-0048949.  
PR 06-JUN-1997; 97US-0048964.  
PR 06-JUN-1997; 97US-0048972.  
PR 06-JUN-1997; 97US-0049020.  
PR 06-JUN-1997; 97US-0049375.  
PR 05-SEP-1997; 97US-0057628.  
PR 05-SEP-1997; 97US-0057635.  
PR 05-SEP-1997; 97US-0057644.  
PR 05-SEP-1997; 97US-0057647.  
  
PI (HUMA-) HUMAN GENOME SCI INC.  
PI Brewer LA, Carter KC, Dillon PJ, Ebner R, Endress GA;  
PI Fan P, Feng P, Fertle AM, Fischer CL, Florence C;  
PI Florence K, Greene JM, Hu J, Kyaw H, Lafleur DW;  
PI Li Y, Moore PA, Ni J, Olsen HS, Rosen CA, Ruben SM;  
PI Shi Y, Soppet DR, Wei Y, Young P, Yu G, Zeng Z;  
XX  
XX WPI; 1999-059865/05.  
DR P-PSDB; W88648, W88934.  
XX  
XX New isolated human genes and the secreted polypeptides they encode -  
PT useful for diagnosis and treatment of e.g. cancers, neurological  
PT disorders, immune diseases, inflammation or blood disorders  
XX  
PS Claim 4; Page 377-378; 772pp; English.

XX CC The invention relates to nucleic acid sequences (V84411 to V84633)  
CC encoding human secreted proteins (W88534 to W88756). The secreted protein  
CC gene sequences are deposited with the ATCC under deposit numbers ATCC  
CC 97979, 97974, 97975, 97976, 97977, 209007, 209008, 209009, 209010,  
CC 209011, 209080, 209081, 209082, 209083, 209084, 209085, 209511. Host  
CC cells comprising recombinant vectors containing the nucleic acid  
CC sequences are used for the recombinant production of the secreted  
CC proteins. The polynucleotide and amino acid sequences are useful for are  
CC useful for preventing, treating or ameliorating medical conditions e.g.  
CC by protein or gene therapy. Pathological conditions can be also diagnosed  
CC by determining the amount of the new polypeptides in a sample or by  
CC determining the presence of mutations in the new polynucleotides.  
CC Specific uses are described for each of the polynucleotides, based on  
CC which tissues they are most highly expressed in, and include developing  
CC products for the diagnosis or treatment of cancer, neurodegenerative  
CC disorders, developmental abnormalities and foetal deficiencies, blood  
CC disorders, tumours, leukemias, diseases of the immune system, autoimmune  
CC diseases, hepatic and renal disease, lymphomas, inflammation, allergies,  
CC ischemic shock, Alzheimer's and cognitive disorders, schizophrenia,  
CC restenosis, prostate diseases, obesity, disorders involving osteoclasts  
CC such as osteoporosis, arthritis or malignancies, diseases of testes, lung  
CC or thymus, digestive/endocrine disorders, infections and AIDS. The  
CC polypeptides are also useful for identifying their binding partners.  
CC The present sequence represents a gene encoding a human secreted protein  
CC (see descriptor line for gene number and clone identification).  
XX CC Sequence 1288 BP: 356 A; 264 C; 295 G; 370 T; 3 other:  
SQ

	Query Match	25.4%	Score 35.8	DB 20	Length 1288
	Best Local Similarity	53.3%	Pred. No. 5.7		
	Matches 73	Conservative 1	Mismatches 63	Indels 0	Gaps 0
QY	1	accgccgagccctgggctcccttgactacccgccggtggcgccgagcgtgtcggtac	60		
Db	153	ggcgccggggccctggggctgcgatggctccgggctccggggcaggcgagctggaggcg	94		
QY	61	cgcgtgagccctgcgggagagcgcaaggcctcgagcgggtccgtgcgggcttcggt	120		
Db	93	gaggagaccgaggagaggaacgtgctgagcgtctggtccggcggtcccgagggccccc	34		
QY	121	cagctgcacccgcgcac	137		
Db	33	aggctgcatcaatgcgc	17		

RESULT	13
A63462	
ID	A63462 standard; DNA; 3035 BP.
XX	
AC	A63462;
XX	
DT	06-MAR-2001 (first entry)
XX	
DE	Streptomyces globisporus C-1027 gene cluster sgcA and sgcB genes.
XX	
KW	Eneidiyne C-1027 biosynthesis gene cluster; apoprotein; chromophore;
KW	sgcA; sgcb; cancer; ds.
OS	Streptomyces globisporus.
XX	
FH	Key
FT	Location/Qualifiers
CDS	101..1099
FT	/*tag= a
FT	/product= "SgcA dNDP-glucose 4,6-dehydratase"
FT	1144..2408
FT	/*tag= b
FT	/product= "SgcB transmembrane efflux protein"
XX	
XX	WO2000040596-A1.
PN	
XX	
PD	13-JUN-2000.

XX	06-JAN-2000; 2000WO-US000445.
XX	06-JAN-1999; 99US-0115434.
XX	05-JAN-2000; 2000US-0477962.
XX	(RECC ) UNIV CALIFORNIA.
XX	Shen B, Liu W, Christenson SD, Standage S;
XX	PI
XX	WPI; 2000-465947/40.
XX	DR
XX	P-PSDB; B13604, B13605.
XX	Isolated nucleic acid comprising a nucleic acid encoding any of C-1027
XX	open reading frames (ORFs) -7 to 42, excluding ORF 9 (cagA), useful for
XX	the production of enediylne C-1027 antitumour antibiotics -
XX	Example 1; Fig 6; 160pp; English.
XX	The present sequence is the sgCA and sgCB genes which form part of the
XX	enediylne C-1027 gene cluster in Streptomyces globisporus. Sgca encodes a
XX	dNDP-glucose 4,6dehydratase enzyme, and sgcb encodes a transmembrane
XX	efflux protein. Enediylne C-1027 is an antibiotic, consisting of an
XX	apoptoin and a non-peptidic chromophore, which acts by damaging DNA.
XX	The sequences within the gene cluster, and the proteins they encode, can
XX	be used in the treatment of cancer, along with antagonists of the
XX	protein.
XX	Sequence 3035 BP; 440 A; 1101 C; 1029 G; 465 T; 0 other;
XX	SQ

Query Match	25.1%	Score 35.4;	DB 21;	Length 3035;
Best Local Similarity	54.1%;	Pred. NO. 6.4;		
Matches 72;	Conservative 0;	Mismatches 61;	Indels 0;	Gaps 0;

  

QY	1	gccgccgagccctgggtctcttgactaccgccgcggtggcgcgcgctgacgtctgtac	60
Db	977	ggctacagagccctgggtctcttgagagcgccctggcgagcggtgaagtgtaccac	1036
QY	61	ccgctgagcgccgtcgggagagcgcaagcctcgagcggtctcgctgcggggttcgt	120
Db	1037	gagaaacgttcgtgggtggcagccctgaagagagcgccgctctggagccgctgcgc	1096
QY	121	cagctgcaccgcgc	133
Db	1097	taacgcagaccac	1109

RESULT	14
A63349	
ID	A63349 standard; DNA; 42000 BP.
XX	
AC	A63349;
XX	
DT	06-MAR-2001 (first entry)
XX	
DE	Streptomyces globisporus C-1027 gene cluster ORF -7-25.
XX	
KW	Eneidyne C-1027 biosynthesis gene cluster; apoprotein; chromophore; cancer; ds.
XX	
OS	Streptomyces globisporus.
XX	
FH	Location/Qualifiers
FT	complement (8...658)
CDS	/tag= a
FT	/product= "ORF -7 protein"
FT	complement (930..1478)
FT	/tag= b
FT	/product= "ORF -6 protein"
FT	complement (1649..2713)
CDS	/tag= c
FT	/product= "ORF -5. protein"
FT	





```
FT      /*tag= ap
FT      /product= "ORF 34 protein"
FT      56026..56880
FT      /*tag= aq
FT      /product= "proline oxidase"
FT      complement (56927..57736)
FT      /*tag= ar
FT      /product= "ORF 36 protein"
FT      complement (57833..58304)
FT      /*tag= as
FT      /product= "ORF 37 protein"
FT      complement (58440..60095)
FT      /*tag= at
FT      /product= "p450 hydroxylase"
FT      complement (60092..60621)
FT      /*tag= au
FT      /product= "ORF 39 protein"
FT      60940..62019
FT      /*tag= av
FT      /product= "ORF 40 protein"
FT      62045..62899
FT      /*tag= aw
FT      /product= "ORF 41 protein"
FT      complement (62787..63164)
FT      /*tag= ax
FT      /product= "ORF 42 protein"
FT      XX
PN      WO200040596-A1.
XX
XX      13-JUL-2000.
XX
XX      06-JAN-2000; 2000WO-US00446.
XX
XX      06-JAN-1999; 99US-0115434.
XX      05-JAN-2000; 2000US-0477962.
XX
XX      (REGC ) UNIV CALIFORNIA.
XX
XX      Shen B, Liu W, Christenson SD, Standage S;
XX      WPI: 2000-465947/40.
XX
XX      P-PSDB; B13554, B13555, B13556, B13557, B13558, B13559, B13560, B13561,
XX      B13562, B13563, B13564, B13565, B13566, B13567, B13568, B13569, B13570,
XX      B13571, B13572, B13573, B13574, B13575, B13576, B13577, B13578, B13579,
XX      B13580, B13581, B13582, B13583, B13584, B13585, B13586, B13587, B13588,
XX      B13589, B13590, B13591, B13592, B13593, B13594, B13595, B13596, B13597,
XX      B13598, B13600, B13601, B13602, B13603, B13606, B13607.
XX
XX      Isolated nucleic acid comprising a nucleic acid encoding any of C-1027
XX      open reading frames (ORFs) -7 to 42, excluding ORF 9 (cagA), useful for
XX      the production of enediyne C-1027 antitumour antibiotics -
XX
XX      Claim 1; Page 78-157; 160pp; English.
XX
XX      The present sequence is the enediyne C-1027 gene cluster from
XX      Streptomyces globisporus. Enediyne C-1027 is an antibiotic, consisting of
XX      an apoprotein and a non-peptidic chromophore, which acts by damaging DNA.
XX      The sequences within the gene cluster, and the proteins they encode, can
XX      be used in the treatment of cancer, along with antagonists of the
XX      protein. Each of the open reading frames is specifically claimed,
XX
XX      Query Match      25.1%; Score 35.4; DB 21; Length 63164;
XX      Best Local Similarity 54.1%; Pred. No. 4.9;
XX      Matches 72; Conservative 0; Mismatches 61; Indels 0; Gaps 0;
XX
QY      1 gcccgcgagcccttgccttcgaactaccgcgcgctggcgccgtgacgctgtcgtaac 60
DB      29466 ggcctacgagcccttgccttcgagcagcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 29525
QY      61 ccgcgtagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 120
DB      29526 gagaaccgttcgtggtgagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 29585
```

```
QY      121 cagctgcacccgc 133
DB      29586 tgacgcgagccac 29598
```

Search completed: June 2, 2001, 00:43:34  
Job time: 2322 sec

